IDENTIFICATION OF DROUGHT-RELATED QUANTITATIVE TRAIT LOCI (QTLs) IN SUGARCANE (*Saccharum* spp.) USING GENIC MARKERS

A Dissertation

by

VIVEK SHARMA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2009

Major Subject: Plant Breeding

IDENTIFICATION OF DROUGHT-RELATED QUANTITATIVE TRAIT LOCI (QTLs) IN SUGARCANE (*Saccharum* spp.) USING GENIC MARKERS

A Dissertation

by

VIVEK SHARMA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Jorge A. Da Silva
	William L. Rooney
Committee Members,	John L. Jifon
	Erik Mirkov
Head of Department,	David Baltensperger

May 2009

Major Subject: Plant Breeding

ABSTRACT

Identification of Drought-Related Quantitative Trait Loci (QTLs) in Sugarcane (*Saccharum* spp.) Using Genic Markers. (May 2009)

Vivek Sharma, B.Sc.; M.Sc., Chaudhary Charan Singh Haryana Agricultural University, Hisar, India

Co-Chairs of Advisory Committee: Dr. Jorge Da Silva
Dr.William L. Rooney

Population based association studies in crops that were established by domestication and early breeding can be a valuable basis for the identification of QTLs. A case control design in a population is an ideal way to identify maximum candidate sites contributing to a complex polygenic trait such as drought. In the current study, marker loci associated with drought related QTLs were identified in sugarcane (*Saccharum* spp), one of the most complex crop genomes, with its polyploid nature (>8), chromosome number (>100) and interspecific origin. The objectives of this investigation were: 1) development of genic markers, which can be used for marker-assisted selection of drought tolerant genotypes of sugarcane. 2) genotypic characterization of sugarcane population at drought related loci using EST-SSR markers. Using 55 microsatellite markers, 56 polymorphisms were scored among 80 modern sugarcane genotypes. Homogeneity of the population was confirmed by determining the distribution of allele frequencies obtained by random genomic microsatellite markers. This analysis was conducted in the STRUCTURE program and the population was divided in 3 subgroups based on the allelic distribution. Phenotypic data to evaluate drought tolerance among the genotypes was collected by measuring chlorophyll content, chlorophyll fluorescence, leaf temperature and leaf relative water content. A generalized linear model in SPSS was used to find association between marker loci and phenotypic data. Markers with significant association ($P \leq 0.001$ level) with the trait were subjected to linear regression to screen the spurious associations. Based on the results, 21 EST-SSR markers and 11

TRAP markers related to drought-defining physiological parameters were considered as genuine associations in this study. Fifty-six polymorphisms produced by 13 EST-SSR primers were used to produce genetic similarity matrix for 80 genotypes. Dendrogram prepared from this genetic similarity matrix will be useful in selecting parents carrying diversity at drought specific loci.

DEDICATION

This work is dedicated to my beloved parents,
Nirmala Devi and Bhuwan Chandra Sharma.

ACKNOWLEDGMENTS

I would like to express my gratitude to those who gave their valuable advice to further my academic development at Texas A&M University.

I want to thank my co-chair Dr. Jorge A. Da Silva for his patience and knowledge.

I was delighted to interact with Dr. William Bill Rooney by attending his class and having him as my co-chair. His insight in quantitative breeding is second to none. Dr. John L. Jifon and Dr. Erik Mirkov deserve special thanks for serving as my dissertation committee members.

Thanks also to my friends, colleagues, and department faculty and staff for making my time at Texas A&M University a great experience. I also want to extend my gratitude to the Texas AgriLife Research Institute, which provided valuable assistance for this research, and to all my colleagues and friends who participated in this study.

A special thanks to Dr. Mona Damaj, Dr. Nael El-Hout and Donna Goodman for their valuable feedback and encouragement. Finally, thanks to my family members for their moral support.

TABLE OF CONTENTS

Page

ABSTRACT……... iii

DEDICATION……... v

ACKNOWLEDGMENTS…… vi

TABLE OF CONTENTS……..……. vii

LIST OF FIGURES……... ix

LIST OF TABLES……. x

CHAPTER

I INTRODUCTION……..... 1

II ASSESSMENT OF DROUGHT TOLERANCE IN SUGARCANE GENOTYPES USING PHYSIOLOGICAL TOOL……... 4

Introduction……. 4
Materials and Methods…… 7
Results……. 9
Discussion……... 12

III DEVELOPMENT OF GENIC MARKERS FOR SELECTION OF DROUGHT TOLERANCE IN SUGARCANE……... 14

Introduction……. 14
Materials and Methods…… 16
Results……. 19
Discussion……... 27

IV ASSESSMENT OF GENETIC DIVERSITY USING DROUGHT-RELATED GENIC MARKERS……..... 33

Introduction……. 33
Materials and Methods……..... 35
Results……. 37

CHAPTER Page

Discussion……... 38

V CONCLUSIONS…….. 43

REFERENCES…….. 44

APPENDIX A……...……. 54

APPENDIX B……...……. 57

APPENDIX C……...……. 59

APPENDIX D……...……. 61

APPENDIX E……...……. 62

APPENDIX F……...……. 63

VITA…….…. 64

LIST OF FIGURES

FIGURE		Page
1	Average monthly rainfall during the period of drought evaluation in two years…….	9
2	Average monthly temperature during the period of drought evaluation in two years…….	10
3	Dendrogram based on Jaccard coefficient and UPGMA clustering method and derived from EST-SSRs data……	39

LIST OF TABLES

TABLE		Page
1	Analysis of variance for chlorophyll fluorescence…….	11
2	Analysis of variance for chlorophyll content…….....	11
3	Analysis of variance for leaf temperature……..	12
4	SSR markers showing marker-trait associations with 2006 data……...	21
5	TRAP markers showing marker-trait associations with 2006 data.....……	22
6	SSR markers showing marker-trait associations with 2008 data……...	23
7	TRAP markers showing marker-trait associations with 2008 data…….	24
8	Determination of population structure using SSR markers unlinked to the trait……	25

CHAPTER I

INTRODUCTION

Sugarcane (*Saccharum* spp. *hybrid*) is a perennial monocotyledonous grass and an important commodity crop adapted to tropical and subtropical regions. The genus *Saccharum* is an important component of the *Andropogoneae* tribe of the grass family *Poaceae*. Sugarcane has acquired, through human selection, a remarkable feature of sequestering carbon into sucrose in the stem. The ability of accumulating high levels of sucrose (~ 0.7 M) in mature internodes makes it distinct among cultivated plants (Moore 1995).

Modern sugarcane cultivars are obtained from interspecific crosses between *S. officinarum* (2*n* = 80 chromosomes) and *S. spontaneum* (2*n* = 40-128 chromosomes), resulting in an asymmetric chromosome transmission that leads to the formation of varieties with different chromosome numbers, generally between 100 and 130 (Price 1965). The highly polyploid and aneuploid nature of the interspecific hybrids allied with the difficulties of controlled hybridization (Hogarth 1987; Silva et al. 1993) and time spent on developing new varieties (from 12 to 15 years) are the main drawbacks in sugarcane breeding programs. Sugarcane offers several other challenges for the molecular geneticists. *S. officinarum* was crossed *with S. spontaneum* (2*n* = 40-128) for its resistance to abiotic and biotic stresses (Roach 1972). Successive backcrosses of the hybrids with *S. officinarum* were carried out to recover the high-sugar producing clones. Most of the modern sugarcane cultivars are derived from this cross. Therefore, genetic variation, which is the key to bring improvement, is limited due to the narrow genetic base (Hogarth 1987) of modern sugarcane cultivars. Furthermore, the large size of the genome makes the search for this minor variation even more cumbersome. Considering these issues, an open population with unrelated individuals has been selected and a

---

This dissertation follows the style and format of Theoretical and Applied Genetics.

linkage disequilibrium based association study was conducted to tag drought related quantitative trait loci (QTLs) in the current study. Since an open population has more diversity, it produces much higher polymorphisms and markers are less likely to be background specific, hence more widely applicable in breeding. Association mapping also gives a higher resolution compared to bi-parental cross because it takes advantage of ancestral recombination (Wei et al. 2006). Using material under routine evaluation reduces the cost of collecting phenotypic data, as special experiments are not required (Jannink and Walsh 2002).

The basic problem with studying a quantitative trait such as drought stress has always been that phenotype of a given genotype reveals little about the genotype itself due to its interaction with the environment. The establishment of quick and efficient screening methods for assessment of drought tolerance in sugarcane (Silva et al. 2007) has shown the way for identification of QTL involved in physiological pathways, contributing to drought tolerance.

Molecular genetic markers are a valuable tool in the studies of complex genomes such as sugarcane (Daugrois et al. 1996). Their incorporation in the selection of economic traits during the early stages of a breeding program, as well as in the choice of the best parents in a cross, may significantly reduce the time for the development of new varieties. These goals can be achieved with the availability of robust polymorphic markers, which co-segregate with economically important traits in sugarcane.

Basically, SSRs are stretches of DNA consisting of repetitive sequences of 2 to 6 base pairs, which are evenly distributed across the genome. They are generally present in non-coding regions of the genome. They are highly polymorphic and flanked by sequences that are sufficiently conserved. Therefore, they are considered as a suitable source for generating primers for PCR amplifications (Powell et al. 1996; Wang et al. 1994). SSR primers are developed based on sequence information, which makes them more reliable and consistent compared to the primers generated at random such as RAPD. This, however, makes their development slow.

The cost of running microsatellite marker analyses is relatively low, but the cost of developing the markers is high, limiting their application to larger commercial crops such as sugarcane. The high cost of developing plant microsatellite libraries, coupled with the low level of enrichment peculiar to sugarcane libraries justify the use of expressed sequence tags (EST)-SSR markers in sugarcane (Silva 2001).

Single-pass partial sequencing of 5' or 3' end of a complimentary DNA (cDNA) clone that represents an mRNA, known as ESTs (Wolfsberg and Landsman 2001), is a fast and efficient approach for sampling the transcribed portion of the genome (Sterky and Lundeberg 2000; Liang et al. 2000). The major advantage of EST-SSR, over genomic SSR (gSSR) markers is that they are economic and time efficient. They are less polymorphic compared to gSSR markers but the polymorphism produced by EST-SSR is more likely to be associated with the trait. Moreover, they are relatively easy to develop.

The objectives of this investigation were 1) Development of genic markers, which can be used for marker-assisted selection of drought tolerant genotypes of sugarcane. 2) Genotypic characterization of sugarcane population at drought related loci using EST-SSR markers.

CHAPTER II

ASSESSMENT OF DROUGHT TOLERANCE IN SUGARCANE GENOTYPES USING PHYSIOLOGICAL TOOLS

Introduction

In the process of identifying marker-trait associations, the phenotypic assessment of the population has to be done, which is later associated with genotypic scoring. The success of an association study is heavily dependent on accurate evaluation of the phenotype of interest. The within population variation observed for genotypes and phenotypes for an association is much greater than that found in most bi-parental mapping populations. Greater variation is preferable while aiming at higher resolution and allele mining, but it can pose problems for accurate evaluation of this variation in a meaningful way in a single environment (Ersoz et al. 2007).

The inherent variation observed in phenotypic trait measurement, when combined with the substantial genetic variation included in some association studies, requires careful experimental design to acquire quality data in multiple environments. Another factor which poses challenge to the accuracy of phenotyping is the nature of the trait. Quantitative traits are difficult to assess since they are highly influenced by the environment. Drought tolerance is by far the most complex trait (Blum 2005) and its definition depends on the phenotyping method used. Many research groups are trying to develop plants that tolerate water stress to face the challenge of increasing water demand for agriculture (Ingram and Bartels 1996). However, transpiration and photosynthesis are intrinsically linked in gas-exchange processes. Biomass accumulation requires light interception by leaves and stomata opening, two processes that are also the main determinants of plant transpiration. The aim is therefore not simply to obtain plants that tolerate cellular water stress, which, under field conditions, is most often avoided by controls at the whole-plant level (e.g. stomata opening, leaf growth and early

senescence) (Tardieu 1996). Rather, the aim is to optimize the trade-off between water use and biomass accumulation (Tardieu 2003).

For complex traits, such as drought tolerance, consideration is often given to the scope for dissection of the primary traits into components that are expected to be simpler to work with compared to the ultimate trait of interest (Ribaut et al. 2004; Campos et al. 2004). Trait dissection is of particular advantage when the component traits themselves have a higher heritability in the mapping population than the ultimate target trait and they have good correlation with the target trait. In the current study, the drought tolerance has been dissected into four physiological parameters namely: chlorophyll content (CC), chlorophyll fluorescence (CF), leaf temperature (LT) and leaf relative water content (LRWC).

Water deficit stress is known to alter a variety of physiological processes such as radiation capture, leaf temperature, stomata conductance, transpiration, electron transport, photosynthesis and respiration, which ultimately determine yield (Qing et al. 2001). The amount of water used by a crop is closely associated with photosynthetic activity, dry matter production and yield in many species (Tollenaar and Aguilera 1992; Qing et al. 2001). However, the maximum photosynthetic potential of crops is seldom reached due to unfavorable environmental factors, including drought.

The degree of limitation of yield by environmental stresses varies even among genotypes within a species (Wolfe et al. 1988; Aguilera et al. 1999). Therefore, the ability to maintain key physiological processes, such as photosynthesis during moderate drought stress, is indicative of the potential to sustain productivity under water shortage. For instance, differences in dry matter accumulation between old and new corn hybrids have been shown to depend on the ability to maintain higher photosynthetic rates after silking for newer hybrids (Tollenaar and Aguilera 1992).

In sugarcane, four distinct growth stages have been characterized, namely: germination, tillering, grand growth and maturity (Gascho and Shih 1983). The tillering and grand growth stages, known as the sugarcane formative phase, have been identified as the critical water demand period (Ramesh 2000), mainly because this is the phase

when 70-80% of cane yield is produced (Singh and Rao 1987). Water relations and photosynthetic responses to water deficit stress during this growth stage could therefore be useful in identifying drought tolerant genotypes.

Although measurements of leaf photosynthesis rate have been shown to be reliable in distinguishing between drought tolerant and susceptible genotypes of some species such as sunflower (Gimenez et al. 1992), gas exchange techniques of assessing photosynthesis are laborious and not practical in crop improvement programs (Earl and Tollenaar 1999). Rong-hua et al. (2006), working with barley, showed indirect and faster methods of measuring photosynthetic activity, such as chlorophyll *a* fluorescence technique, particularly the maximum photochemical efficiency of photosystem II – PSII (which can be assessed via the variable fluorescence (*F*v) to maximum chlorophyll *a* fluorescence (*F*m) ratio). Estimated chlorophyll content (SPAD index), can be as effective as the more time-consuming gas exchange techniques in revealing differences between drought tolerant and susceptible genotypes. Other physiological parameters such as LT and LRWC are also very responsive to drought stress and have been shown to correlate well with drought tolerance (Jamaux et al. 1997; Altinkut et al. 2001; Colom and Vazzana 2003). The reliability of these parameters to distinguish between stress tolerant and susceptible genotypes in sugarcane has been proven by Silva et al. (2007).

The utilization of different QTL information dramatically influences marker assisted selection (MAS) efficiency. QTLs involved in MAS when evaluated in a single environment, leads to QTL × environment (*QE*) interactions, which usually reduce general response across environments, and the reduction in the cumulative general response is a function of the proportion of *QE* interactions for the trait studied. However, MAS using QTL information evaluated in multiple environments not only yields higher general response, but also have the additional advantage of robustness in *QE* interactions. The total response achieved by MAS in a specific environment depends largely on the total heritability of traits and is slightly subject to relative changes between general heritability and *GE* interaction heritability. Considering these factors, several studies have concluded that plant breeders should be cautious to utilize QTL

information from only one environment and perform breeding studies in another (Liu et al 2003).

The objective of this study was to observe the two year phenotypic data in terms of variance components. It was important to determine if the effect of treatment was significant in year 2008 data, since period of treatment was short. Also determining the variances in two year data was important to see if a combined analysis can be performed on two years data.

Materials and Methods

The current study was conducted at Weslaco (26° 12' N, 97° 57' W, elevation 18.90m), Texas, USA, during the 2005-2006 and 2007-2008 growing seasons in a commercial field with a sandy clay loam soil type. The experiment was designed as a complete block within a four-factor factorial, where the first factor was year, the second factor included two irrigation levels (wet and dry), the third factor contained eighty genotypes and the fourth factor was represented by three evaluation dates (2006 and 2008). The data was collected from four replicates.

Eighty sugarcane genotypes analyzed in this study were selected from collections maintained by the Texas AgriLife Research Center in Weslaco, Texas. This population included commercial cultivars as well as elite clones from diverse pedigrees. Plants were planted 1.5m apart from each other in 3m long rows. Planting was done on November 14th 2005. The data was collected on plant cane in 2006 and on second ratoon in 2008.

The plants were grown without any water stress until grand growth phase and the two irrigation treatments were initiated at the onset of grand growth phase. The well-watered side of the field was irrigated at 50% depletion of available soil moisture (DASM), whereas dry plots were irrigated at 80% DASM. Soil moisture depletion was monitored periodically with neutron probes, which were placed in the ground at four different levels (1, 2, 3 and 4 feet) for the two irrigation treatments (well-watered and drought).

Physiological parameters were measured three times during the study in 2006 at 0, 45 and 90 days after treatment (DAT) after the start of irrigation treatment. In 2008 only two observations at 0 and 45 DAT could be made due to flooding caused by hurricane Dolly. Measurements were taken on cloudless days between 0900 h and 1500 h. Chlorophyll *a* fluorescence characteristics were measured on intact leaves using a pulse amplitude modulation flurometer (Model OS5-FL, Opti-Sciences, Tyngsboro, MA, USA), and used to estimate the extent of drought-induced photo-inhibition. At least four leaves were dark adapted for 30 min using leaf clips (FL-DC, Opti-Science) before fluorescence measurements. The chlorophyll fluorescence (Fv/Fm) parameter was determined following the procedure of Maxwell & Johnson (2000).

Leaf chlorophyll content (SPAD index) was estimated using a SPAD-502 chlorophyll meter (Minolta Corp., Ramsey, NJ, USA). This method is nondestructive and the readings in the chlorophyll meter correlates well with leaf chlorophyll content (Markwell et al. 1995). It has been used as a promising tool for rapid and nondestructive screening for drought tolerance in sugarcane (Silva et al. 2007). Five measurements were taken and averaged from different plants in each plot.

Leaf temperature (LT) readings were collected using a hand-held infrared thermometer (Model OS530HR, Omega Engineering Inc., Stanford, CT, USA). Leaf emissivity was set at 0.95 during observations. During each LT measurement, the natural leaf orientation with respect to the sun was maintained to avoid shade effects.

Leaf relative water content (LRWC) was determined following the method of Matin et al. (1989). Two leaf disks (1.3 cm diameter each) per genotype were collected with a cork borer from the same leaves used for *Fv/Fm*, SPAD index and LT measurements. Disks were immediately sealed in glass vials and quickly transported to the laboratory in an ice-cooled chest. Leaf disk fresh weights were determined within 2 h after excision. The turgid weight was obtained after rehydration in de-ionized water for 24 h at room temperature. After re-hydration, leaves were quickly and carefully blotted dry with lint-free tissue paper before determining turgid weight. Dry weights were recorded after oven-drying leaf samples for 48 h at 80ºC.

The analysis of variance (ANOVA) was performed on 0 DAT and 45 DAT observations to determine the variance components of CC, CF and LT. Observations made on 90 DAT in 2006 were ignored, in order to determine the year effect in this balanced data. Block and irrigation treatment were considered as fixed variables while genotype and environment were considered as random variable. Evaluation dates were considered as repeated observations in the analysis. The ANOVA was conducted on both 2006 and 2008 data individually as well as collectively to find the year effect in total phenotypic variation. PROC ANOVA was used in SAS for these analyses.

Results

Significant effects ($P<0.01$) were observed for different interactions on several physiological parameters, and data of year 2006 and 2008 varied significantly ($P<0.01$). Monthly average rainfall during the data collection period (April-August) for these two years was 126.5 mm and 709.42 mm respectively (Fig. 1). Average air temperatures during the period of evaluation (November 2005 to August 2006 and November 2007 to August 2008) ranged from 25.5 to 29.45 °C and 30.5 to 34.3 °C respectively (Fig. 2).

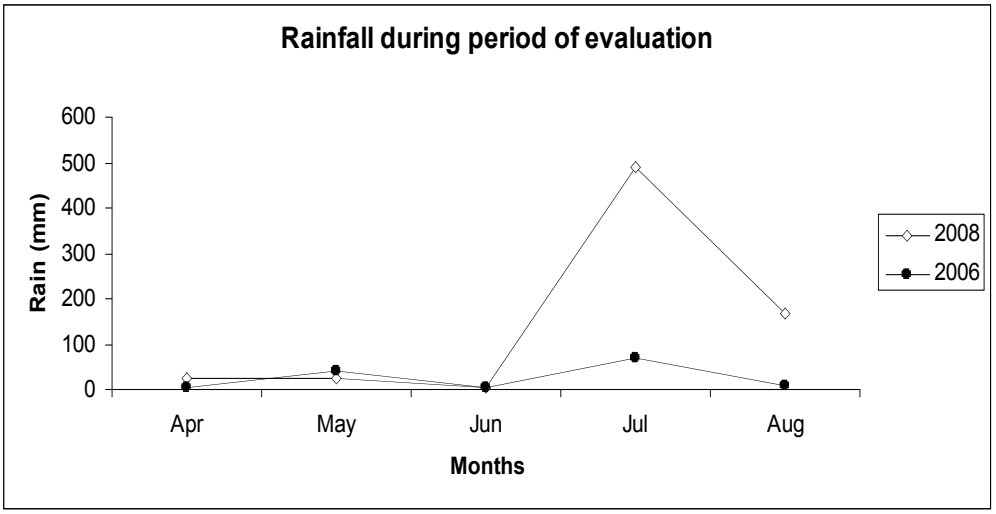

Fig. 1 Average monthly rainfall during the period of drought evaluation in two years

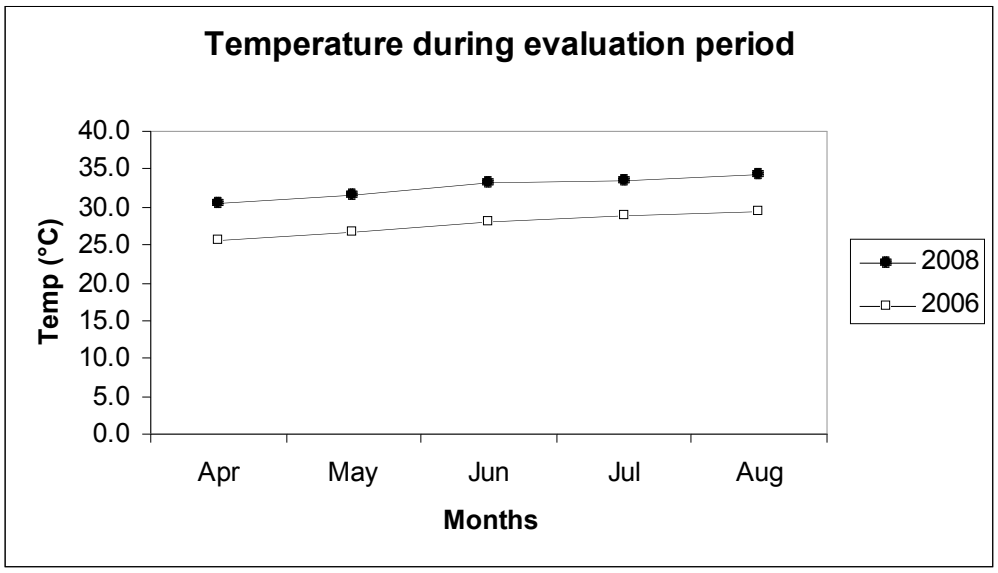

Fig. 2 Average monthly temperature during the period of drought evaluation in two years

Significant genotype by treatment (G x T) and genotype by treatment by evaluation date (G x T x ED) interactions were observed for photosystem II (PSII) photochemical efficiency (*Fv/Fm*) measurements in year 2006. In year 2008 data, G x T interaction was not significant but the remaining components had significant effects. When data of years 2006 and 2008 were analyzed together with the year as a component in total variation and subjected to ANOVA, non-significant effect was obtained for G, ED and G x T (Table 1).

Table 1 Analysis of variance for chlorophyll fluorescence

Effect	F value		
	2006	2008	2006-2008
Block (B)	7.86**	53.02**	25.56**
Year (Y)	-	-	14.73**
Treatment (T)	19.54**	58.27**	37.37**
Genotype (G)	3.07**	2.57**	1.02 ^{ns}
Evaluation Date (ED)	7.62**	103.81**	1.42 ^{ns}
G*T	2.43**	0.94 ^{ns}	1.04 ^{ns}
Y*T	-	-	22.30**
G*T*ED	6.30**	1.72**	1.34**

* = P<0.05, ** = P<0.001, ns = non significant

The one-way interactions in variance components on chlorophyll content (SPAD index) were significant in year 2006. However, the interaction effect of G x T and G x T x ED was non-significant. The G x T x ED interaction affected the SPAD index significantly in year 2008 (Table 2). Significant Y x T and G x T x ED interactions were observed in 2006-2008 collective analysis.

Table 2 Analysis of variance for chlorophyll content

Effect	F value		
	2006	2008	2006-2008
Block (B)	1.32**	40.76**	2.40**
Year (Y)	-	-	33.83**
Treatment (T)	3.77**	161.89**	13.98**
Genotype (G)	3.36**	13.14**	7.58**
Evaluation Date (ED)	10.05**	426.89**	147.53**
G*T	0.21^{ns}	1.14 ^{ns}	0.77 ^{ns}
Y*T	-	-	4.52**
G*T*ED	0.88 ^{ns}	2.81**	2.16**

* = P<0.05, ** = P<0.001, ns = non significant

The main effect of treatment was not significant for leaf temperature in year 2006. Significant G x T x E interactions were observed. In year 2008 the genotype and

evaluation date as main effects were not significant. G x T and G x T x ED interactions were not observed. Y x T significantly affected the leaf temperature in the combined analysis of 2006-2008 (Table 3).

Table 3 Analysis of variance for leaf temperature

Effect	F value		
	2006	2008	2006-2008
Block (B)	2.28**	125.33**	3.23**
Year (Y)	-	-	4.43**
Treatment (T)	0.44 ^{ns}	11.77**	3.98*
Genotype (G)	2.24**	1.11 ^{ns}	1.00 ^{ns}
Evaluation Date (ED)	13.67**	0.14 ^{ns}	1.23 ^{ns}
G*T	0.68 ^{ns}	0.23 ^{ns}	0.42 ^{ns}
Y*T	-	-	2.66*
G*T*ED	2.25**	0.59 ^{ns}	0.74 ^{ns}

* = P<0.05, ** = P<0.001, ns = non significant

Collecting leaf disks was a time consuming process. Therefore, measurements for LRWC were made only on one block in both treatments. Consequently, ANOVA could not be conducted on LRWC.

Discussion

Crop cultivars are released based on a target environment, and G x E interactions create problems in estimating the exact yield of the genotype. Similarly, QTL x E interaction affects the stability of the marker. Multiple environment analysis is therefore essential to determine the stability of the genotype or QTL. Environment normally includes season and location. Current study was conducted at a single location for two seasons to determine the stability of QTLs through the years.

Even though the effect of treatment was found significant (P<0.01) for all parameters (CC, CF, LT), the year as a main effect and its interaction with treatment were found significant (P≤0.05) in CF, CC and LT when data of two years were pooled

(Table 1, 2 and 3). This clearly indicates that there is a significant difference between the data collected during the two years. In addition, the main effect of the genotype on CF and LT, which was significant ($P \leq 0.05$) in each year analysis, was observed as being non-significant when the two years data were pooled (Table 1 and 3). This indicates that the major variation in the collective data was due to year.

This variation in two year data could be due to several factors. One major factor was the difference in cane stage. Since in year 2006, observations were made on plant cane, while in year 2008, the observations were made on second year ratoon, the variation in two years is not surprising. The physiology of the sugarcane plant during these two stages is different. It is very likely that in ratoon stage tolerant plant show more tolerance and susceptible plant show more susceptibility.

Considering the results of analysis of variance, it is evident that the excess rain in year 2008 due to hurricane Dolly affected the drought treatment in the month of July (Fig. 1). Therefore, ANOVA was conducted only on initial two observations (0 DAT and 45 DAT). But study of average monthly rainfall may not be enough to study the impact of rainfall on drought treatment. Distribution of rainfall with in a month may cause variation in drought treatment. Average temperature was also higher in year 2008 as compared to year 2006 (Fig. 2).

The possibility of conducting analysis collectively on two years data was explored. But the variances in the two years were not equal. Therefore, the two years data was kept separate for marker-trait analysis with the idea that QTLs identified as consistent in both the years will be considered as robust and free from QTL environment interactions.

CHAPTER III

DEVELOPMENT OF GENIC MARKERS FOR SELECTION OF DROUGHT TOLERANCE IN SUGARCANE

Introduction

Progress in molecular breeding of sugarcane is slow because of its complex genome (Hogarth 1987). Sugarcane cultivars used in breeding programs are interspecific hybrids between the domesticated species *Saccharum officinarum* and the wild relative *S. spontaneum*, followed by repeated backcrossing to *S. officinarum*. The interspecific origin, the high ploidy level (>8) and the high chromosome number (>100) of these cultivars, with 80% *S. officinarum* and 10% *S. spontaneum* (Grivet et al. 1996), contribute to the genetic complexity of sugarcane.

Apart from this genetic complexity, modern sugarcane cultivars have a narrow genetic base resulting from a single cross between *S. officinarum* and *S. spontaneum* (Srinivasan et al. 1987). This narrow genetic base prevents the identification of a sufficient number of polymorphic markers, limiting the power of modern mapping techniques for crop improvement.

Association mapping and linkage disequilibrium (LD) mapping based on unrelated populations is an ideal approach to explore allele specific variation. This strategy takes advantage of ancestral recombination. It does not require a specific mapping population and enables gene level mapping on non-model organisms (Nordborg and Tavare 2002; Risch and Merikangas 1996). It is especially advantageous for sugarcane because its genome is not completely sequenced. Moreover, developing recombinant inbred lines is very difficult in sugarcane considering its long life cycle, vegetative reproduction and out crossing sexual reproduction. LD is high in sugarcane, and fewer markers are therefore needed to saturate the genome as compared to other crops. The discovery of polymorphic loci should be very useful in manipulating breeding germplasm in elite x elite crosses.

Quantitative traits have been a major area of genetic study for over a century because they share a common feature of natural variation in populations of all eukaryotes. Drought is one such trait and it negatively influences survival, biomass production and crop yield (Pennisi 2008). Drought stress alters a majority of physiological processes that ultimately determines yield, such as radiation capture, leaf temperature, stomata conductance, transpiration, electron transport, photosynthesis and respiration (Qing et al. 2001; Silva et al. 2007). Understanding the mechanisms of drought tolerance and breeding for drought-tolerant crop plants has been the major goal of plant biologists and crop breeders. However, drought is considered as a multigenic quantitative trait and it is a challenge to understand its molecular basis and to manipulate drought tolerance. In addition, the effect of individual genes on phenotype is generally very small and inconsistent across the environment. Increasing the tolerance of crop for drought stress would be the most economical approach to improve crop productivity and reduce agricultural use of water resources. However, drought tolerance is difficult to manage for molecular geneticists mainly due to the limited awareness of the specific traits that are linked to it.

One way to overcome this problem is to identify fragments associated with factors that regulate the expression of several stress-related genes. Simple sequence repeats (SSRs) derived from expression sequence tags (EST) can be associated with genes of known function and used as functional SSR markers, thus tagging genes of interest in a more efficient manner (Hackauf and Wehling 2002). Thirteen EST-SSR primers have been used in the present study (Appendix E). These primers were based on candidate genes involved in stress-responsive pathways. Target region amplification polymorphism (TRAP) primers have also been designed in this study, based on dehydration responsive element binding (DREBs) genes, since no microsatellite sequences were found in their sequences. DREBs are transcription factors that bind to drought responsive *cis*-acting elements. They belong to a family of transcription factors known as ethylene responsive factors (ERF) and consist of two subclasses, i.e. DREB1 and DREB2 that are induced by cold and dehydration, respectively (Agarwal et al.

2006). DREBs are the major genes involved in triggering the dehydration stress-responsive pathway.

The objective of this study was to investigate if association mapping in sugarcane could be useful in identifying EST based markers for the selection of drought tolerant genotypes. The effect of population structure on identifying marker-trait associations was also assessed because, if present, such association would be of limited value for marker-assisted selection in breeding programs. Several EST-SSR and TRAP markers were developed as useful tools to assist sugarcane breeding programs in identifying drought-responsive genes without phenotypic evaluation.

Materials and Methods

Population selection and phenotyping

A panel of 80 genotypes (Table 1) was selected from a sugarcane collection maintained by the AgriLife Research Center at Weslaco. These clones were derived from diverse pedigrees and do not form any pre-designed mapping population. Some clones, bred and selected outside the United States were also included, and tested in our local environment. They may further be used as a source of genetic diversity.

Phenotyping of the population involved the measurement of quantitative physiological traits, which have been reported as quick and effective tools to evaluate drought tolerance in large plant populations (Silva et al. 2007). These physiological parameters were chlorophyll content (CC), chlorophyll fluorescence (CF), leaf temperature (LT) and relative water content (LRWC).

Data mining

EST-SSR markers were derived from cluster consensus sequences of stress-related ESTs obtained from the Sugarcane Expressed Sequence Tag database (SUCEST at /http://sucest.lbi.dcc.unicamp.br/en/) as described by Pinto et al. (2004). Fifty-five sets of EST-SSR markers were designed using the Primer3 software (Whitehead Institute for Biological Research at http://www.genome.wi.mit.edu/) and synthesized by MWG

Biotech AG. EST sequences for DREB genes were searched at TIGR-SOGI database (http://www.tigr.org) and forward PCR primers were designed from EST sequences of four of these genes using the software Primer3 (Table 2) developed by the Whitehead Institute for Biological Research and available at http://www.genome.wi.mit.edu/. The parameters used for the design of the primers were based on Hu and Vick (2003).

Molecular analysis

Plant DNA was extracted from 50 mg of fresh leaf tissue using the CTAB method (Doyle and Doyle 1990). Arbitrary primers were synthesized with an infrared modification (IRDye700 or IRDye800) to allow the visualization of the PCR products. PCR reactions for target region amplification polymorphism (TRAP) markers were performed in an MJ PTC-100 thermocycler (MJ Research) under the same conditions as described in Hu and Vick (2003), and amplified fragments were visualized in a Li-COR 4300 DNA analyzer (Li-COR, Lincoln, NE) according to the manufacturer's recommendations. Arbitrary primers 2 and 3 from the study conducted by Hu and Vick were used in this study. PCR reaction for SSR primers contained 15ng of DNA template, 2ρmol of each reverse and M13 (-21) primers and 0.4ρmol of the forward primer, 0.2mM dNTPs, 1x PCR buffer and 1 unit of *Taq* DNA polymerase (Promega, Madison, WI) in a final volume of 10μl. PCR conditions were as follows: Preheating, 94 ºC for 5 min, and cycling, 30 cycles of 94ºC (30s)/56ºC (45s)/72ºC (45s), followed by 8 cycles of 94ºC (30s)/53ºC (45s)/ 72º C (45s), and a final extension at 72ºC for 10 min.

Data analysis

Genotypes were grouped using the program STRUCTURE v2 (Pritchard et al. 2000). This analysis was conducted on 32 markers generated from random genomic SSR markers. STRUCTURE program used the marker data to cluster genotypes into groups based on similarity on overall marker profiles.

The Generalized linear model of SPSS was used to seek the initial evidence for association between presence or absence of the marker and each of the physiological

parameters. Threshold values for *t*-test were determined assuming normal distribution of trait values. The generalized linear model expands the general linear model so that the dependent variable is linearly related to the factors and covariates via a specified link function. This model is a regression model containing only categorical independent variables. Regression analysis was performed using the SPSS program, with the genotype information generated by each marker as the indicator variable, which received the value 1 or 0 for presence or absence, respectively. Each indicator variable was regressed against the physiological parameters which were studied on the population in Chapter II.

It is recognized that marker-trait associations in the population used could be spurious due to embedded population structure (Jannink and Walsh 2002) or random variation (type 1 statistical errors). Markers that were significant at stringent threshold of $P<0.001$, based on individual comparison-wise tests, were therefore selected for permutation testing (Churchill and Doerge 1994). In this approach, the selected markers were used at once in the model and their relative main effect was determined. Markers that were significant at $P<0.05$ were selected.

To test for the contribution of marker-trait associations due to population structure, the following model was fitted for the markers, which were significantly associated with the traits:

Physiological parameter = group + marker + group*marker + residual

where groups were determined by STRUCTURE through a best fit approach. The analysis was carried out in SPSS using the general linear model. The presence of significant marker*group interaction ($P<0.05$) would indicate that the effects of the marker differed depending on the group (ancestral background). Therefore, such markers were flagged for further validation. Absence of marker*group interaction, on the other hand, would indicate that there is no influence of ancestral background on marker-trait association. Further validation of markers with marker*group interaction was done by

observing marker within group variation. If marker within group variation is high then marker*group interaction can not be concluded as the effect of population structure.

Results

Phenotyping evaluation involving four physiological parameters, i.e. CC, CF, LT and LRWC was performed in the growing season of 2006 and 2008 at Hiller farm, Weslaco. The Crop was planted in 2006 and the same observations were made in 2008 on the second ratoon crop. Genotypes were evaluated in a case control study. In order to capture maximum variation in a short period of time, the study was conducted during the formative stage of growth. Results of these analyses showed large amount of variation (Appendix 2 and 3) which justified the level of polymorphism obtained.

Association between markers and physiological parameters

The *t*-test for association between marker and trait was conducted using the generalized linear model in SPSS with normal distribution and identity as link function. For the EST-SSR markers, twenty-one out of 56 polymorphic markers were found to be significantly associated with either or both of the two physiological parameters (CC and LRWC) at $P<0.001$ (Table 4). For the TRAP markers, eleven polymorphic markers showed association with either or both of CC and LRWC at $P<0.001$ (Table 5). Twenty-one marker-trait associations among EST-SSRs and the 11 from TRAPs were much higher than 0.056 and 0.014 respectively, which would be expected to give *t*-values greater than the $P<0.001$ threshold level by random chance. Seven EST-SSR markers significantly associated with more than one parameter at $P<0.001$ were SSR 9a, SSR 9b, SSR 048a, SSR 80a, SSR 80b, SSR 230c and SSR 230e. Similarly, five TRAP markers were significant for more than one physiological parameter ($P<0.001$). These markers were TRAP 312c, TRAP 312d, TRAP 313c, TRAP 423b and TRAP 423d. Fourteen markers associated with CC contributed 32.5% to the total variation. However, 17 markers associated with LRWC explained only 11.6% of the total variation. Similarly, 9 TRAP

markers associated with CC explained 9.7%. Fourteen TRAP markers contributed 15.4% to the total variation in LRWC.

Markers significant at P<0.001 were validated by collecting the phenotypic data in year 2008 on the ratoon crop. In spite of big phenotypic variation due to difference in weather and crop stage, twelve SSR markers and all TRAP markers showed consistent association with the phenotypes (Tables 6 and 7). Several markers which were significantly associated with the phenotype in year 2006 did not show any association in year 2008. However, only SSR 9d was a new marker-trait association in 2008 data.

In order to identify the markers with major main effect over the trait, the experiment-wise permutation testing was conducted with 2006 data. Markers which were significant at P<0.001 in the *t*-test (Tables 4, 5, 6 and 7) were included in models with CC, LT and LRWC as dependent variables. SSR 9b, SSR 80a, SSR 230c markers were found significantly associated (P<0.05) with CC. No SSR marker was found associated with LRWC. Among TRAP markers, TRAP 312a was found associated with CC (P<0.05) whereas TRAP 312c, TRAP 312e and TRAP 423d were found associated with LRWC (P<0.05).

Table 4 SSR markers showing marker-trait associations with 2006 data

Marker level	Trait	Probability (P≤0.001)	Group* marker (P≤0.05)	Marker within group (P≤0.05)
SSR 4	CC[a]	0.001	0.043	0.204
SSR 9a	CC	0.000	0.023	0.040
	LRWC[b]	0.000	0.222	0.222
SSR 9b	CC	0.000	0.931	0.066
	LRWC	0.000	0.172	0.304
SSR 9c	LRWC	0.000	0.423	0.200
SSR 9e	LRWC	0.000	0.225	0.197
SSR 668a	LRWC	0.000	0.222	0.473
SSR 668d	LRWC	0.000	0.605	0.103
SSR 048a	CC	0.001	0.503	0.787
	LRWC	0.001	0.835	0.468
SSR 048b	LRWC	0.000	0.199	0.107
SSR 80a	CC	0.000	0.015	0.097
	LRWC	0.001	0.022	0.013
SSR 80b	CC	0.000	0.006	0.804
	LRWC	0.000	0.528	1.161
SSR 80e	CC	0.000	0.050	0.545
	LRWC	0.000	0.580	0.989
SSR 80f	CC	0.000	0.391	0.610
SSR 12c	CC	0.000	0.001	0.317
SSR 50b	CC	0.000	0.921	0.243
	LRWC	0.000	0.714	0.181
SSR 23a	CC	0.000	0.577	0.310
	LRWC	0.000	0.701	0.238
SSR 23b	LRWC	0.000	0.877	0.579
SSR 230a	LRWC	0.000	0.871	0.311
SSR 230c	CC	0.000	0.846	0.040
	LRWC	0.000	0.211	0.119
SSR 230d	CC	0.000	0.459	0.293
SSR 230e	CC	0.001	0.054	0.118
	LRWC	0.000	0.056	0.342

a- chlorophyll content, b- leaf relative water content

Table 5 TRAP markers showing marker-trait associations with 2006 data

Marker level	Trait	Probability ($P \leq 0.001$)	Group*marker ($P \leq 0.05$)	Marker within group ($P \leq 0.05$)
TRAP 312c	CC[a]	0.000	0.939	0.386
	LRWC[b]	0.000	0.071	0.003
TRAP 312d	CC	0.000	0.126	0.448
	LRWC	0.001	0.003	0.056
TRAP 312e	LRWC	0.000	0.372	0.025
TRAP 312f	CC	0.000	0.688	0.056
TRAP 313a	LRWC	0.000	0.715	0.131
TRAP 313b	LRWC	0.000	0.010	0.107
TRAP 313c	CC	0.000	0.354	0.212
	LRWC	0.000	0.908	0.323
TRAP 423b	CC	0.001	0.098	0.362
	LRWC	0.001	0.235	0.194
TRAP 423c	LRWC	0.000	0.629	0.418
TRAP 423d	CC	0.000	0.066	0.124
	LRWC	0.000	0.433	0.043
TRAP 423e	LRWC	0.000	0.582	0.247

a- chlorophyll content, b- leaf relative water content

Table 6 SSR markers showing marker-trait associations with 2008 data

Marker level	Trait	Probability ($P \leq 0.001$)	Group*marker ($P \leq 0.05$)	Marker within group ($P \leq 0.05$)
SSR 9a	C.C[a]	0.000	0.558	0.047
	LRWC[b]	0.000	0.507	0.593
SSR 9b	C.C	0.001	0.342	0.172
	LRWC	0.000	0.315	0.581
SSR 9c	LRWC	0.000	0.080	0.058
SSR 9d	C.C	0.000	0.663	0.278
SSR 9e	C.C	0.000	0.735	0.377
	LRWC	0.000	0.246	0.472
SSR 668a	LRWC	0.000	0.610	0.253
SSR 048b	LRWC	0.000	0.547	0.387
SSR 80a	C.C	0.000	0.576	0.478
	LRWC	0.000	0.661	0.985
SSR 80b	C.C	0.000	0.592	0.059
	LRWC	0.000	0.366	0.262
SSR 230a	LRWC	0.001	0.432	0.419
SSR 230c	C.C	0.000	0.835	0.470
SSR 230d	C.C	0.000	0.276	0.202
SSR 230e	C.C	0.000	0.303	0.726
	LRWC	0.000	0.542	0.157

a- chlorophyll content, b- leaf relative water content

Table 7 TRAP markers showing marker-trait associations with 2008 data

Marker level	Trait	Probability (P≤0.001)	Group*marker (P≤0.05)	Marker within group (P≤0.05)
TRAP 312a	CC[a]	0.000	0.643	0.048
TRAP 312b	CC	0.000	0.109	0.892
	LRWC[b]	0.000	0.015	0.025
TRAP 312c	CC	0.000	0.927	0.132
	LRWC	0.000	0.421	0.813
TRAP 312d	CC	0.000	0.213	0.726
	LRWC	0.001	0.334	0.351
TRAP 312e	LRWC	0.000	0.614	0.605
TRAP 312f	LRWC	0.000	0.159	0.726
TRAP 313a	CC	0.000	0.322	0.157
	LRWC	0.000	0.995	0.191
TRAP 313b	LT[c]	0.000	0.071	0.043
	LRWC	0.000	0.752	0.371
TRAP 313c	CC	0.000	0.638	0.445
	LRWC	0.000	0.550	0.217
TRAP 423a	LT	0.000	0.952	0.419
	LRWC	0.000	0.066	0.143
TRAP 423b	LRWC	0.000	0.389	0.104

a- chlorophyll content, b- leaf relative water content, c- Leaf temperature

Effects of population structure on associations between markers and phenotypes

EST-SSR markers functionally unlinked to CC and CF were initially used to detect the spurious associations due to population structure (Table 8). Two EST-SSR markers were selected and their association with phenotypes was evaluated using the generalized linear model in SPSS. None of these markers were found significantly associated with CC and CF at P<0.001 level.

Table 8 Determination of population structure using SSR markers unlinked to the trait

EST Annotation	Markers	Trait	Probability (P≤0.001)
Auxin-independent growth promoter	SSR-924	CC[a]	0.024
		CF[b]	0.993
Pathogenesis related protein PRMS precursor	SSR-52	CC	0.073
		CF	0.989

a-chlorophyll content, b-chlorophyll fluorescence

However, considering the size of sugarcane genome, two EST-SSR markers unlinked to the trait of interest were not enough to determine the structure in the population. Therefore, another approach was used in which 32 random genomic SSR markers were used in STRUCTURE program to determine the groups in the population. The number of groups (K) in the population was determined according to Wei et al. (2006). The K value was arbitrarily assigned based on an estimate of a "goodness of fit" parameter ln P(X/K), which is an informal pointer provided by STRUCTURE for the number of subpopulations (Pritchard et al. 2000). The value of ln P(X/K) increases with increasing group number and reaches a plateau when a assigned number for group matches with underlying groups in allele frequency. A plateau is indicated when the difference in ln P (X/K) values for two consecutive K values is within 5 to 10 for small to moderate size data sets. In the current study, differences of ln P(X/K) was 8 when K was 1 and 2 but it was reduced to 4 between K = 2 and 3. Under stringent conditions, three groups were considered for analysis of population structure in this study.

The presence of distinct structural features without strong grouping discontinuities is typical of parent collections used in sugarcane breeding programs. In this situation, marker-trait association could easily arise from an uneven contribution of particular parents or ancestors, without necessarily being due to physical linkage of markers and QTL affecting the phenotype of interest (Wei et al. 2006).

When population structure at the three group levels was taken into account, five SSR markers (SSR 9a, SSR 12c, SSR 80a, SSR 80b and SSR 80e) showed significant interaction ($P<0.05$) with the group in year 2006. However, SSR 9a and SSR 80a also showed significant marker effect. In year 2008, no marker*group interaction was detected. The markers, TRAP 312D, TRAP 312F and TRAP 313B, showed significant interaction with the group variable at $P<0.05$ in year 2006, and the marker TRAP 312B showed significant interaction in year 2008. None of them showed significant marker effect. Therefore, these markers should not be used for marker-assisted selection studies aiming at drought stress tolerance in this sugarcane population.

Markers which were significant ($P<0.001$) based on t-test were further divided in to four categories. Those which did not show significant ($P<0.05$) interaction with the group and showed significant ($P<0.05$) within group variation were categorized as best markers. Markers which showed significant ($P<0.05$) interaction with the group but also showed significant ($P<0.05$) within group variation were categorized as second category markers. Markers which did not show significant ($P<0.05$) interaction with the group and also did not show significant ($P<0.05$) within group variation were categorized as third category markers. Markers which showed significant ($P<0.05$) interaction with the group and also did not show significant ($P<0.05$) within group variation were categorized as spurious associations and rejected.

SSR 9a, SSR 80a, SSR230c and TRAP 312a showed consistent association with CC in different environments and different statistical procedures. Similarly, TRAP 312c, TRAP 312e and TRAP 423d were found consistently associated with LRWC under different environments and different statistical procedures. These markers were further analyzed with yield components such as stalk diameter, stalk number, stalk height and stalk weight, which are good indicators of drought tolerance (Silva et al. 2008). Associations were detected using simple linear regression. SSR 9a was found significantly associated with stalk diameter ($P<0.05$).

Discussion

The large number of markers identified in this study, reflect the usefulness of the mapping of association approach for marker discovery within a breeding population. Population stratification can lead to a high incidence of type 1 errors in detecting marker-trait associations within germplasm not derived from a bi-parental cross (Pritchard and Rosenberg 1999). Therefore, evidence of sub-structure was taken into account while finding associations in this study. These results are consistent with previous findings that modern sugarcane has high linkage disequilibrium (Janoo et al. 1999) and short breeding history (Raboin LM et al. 2008).

Some of these markers may represent allelic diversity at the same locus. Dominantly scored SSR fragments are anonymous in nature, so it is not possible to postulate any potential allelic relationships between them without a genetic map where homology relationships between linkage groups can be identified.

Another explanation for the larger number of significant markers identified is that it is likely that some markers are linked on the same haplotype. Some quantitative trait alleles (QTAs) may have several markers linked to them. The number of functional alleles will be less than the number of marker-trait associations in that case. But, without a map showing linkage arrangements between markers, these relationships will remain unknown (Butterfield et al. 2004).

The number of markers used in this study is not large enough for extensive genome coverage in sugarcane, but due to extensive LD these markers may still be able to detect many marker-trait associations.

Population structure is the major issue in association studies. Use of markers unlinked to the trait of interest is one of the methods suggested to detect the effect of structure in the population (Pritchard et al 2000). By choosing drought tolerance as trait of interest, it was hard to identify genes unlinked to the trait because the whole plant-metabolism is affected by water deficit. Population structure was therefore detected specifically with the genes, which are not expected to be functionally involved in CC

and CF pathways (Table 5). However, their physical linkage with the trait can not be ruled out in the absence of information about their location on the chromosome.

The use of evenly distributed random genomic markers has been suggested by Pritchard et al (2000) in order to detect the subgroups in the population based on differences in the allele frequencies. The number of markers required for complete coverage of the genome depends upon the size of the genome and distribution of markers. The use of as low as 36 markers has been reported in wheat (Bresghello and Sorrells 2006), which is comparable to sugarcane in terms of genome size. Detection of subgroups in the population can be accomplished by using STRUCTURE program. It is difficult to determine the exact number of sub-groups, due to the absence of strong grouping discontinuities. The population has been divided into 3 sub-groups based on the Q value from STRUCTURE. It should be noted that resolution of groups and composition in breeding population is an arbitrary process based on "goodness of fit". This area needs further investigation.

The drought trait selected for our association study is multigenic and quantitative and poses a problem in the identification of loci that have a major impact on the phenotype. In spite of large LD in sugarcane (Janoo et al. 1994, Raboin et al. 2008) and measuring drought tolerance in terms of physiological parameters as general as LRWC, several polymorphic markers did not show association with any of CC, CF, LT, LRWC. There could be several reasons for this. These markers could be gene targeted markers (GTM), which is a category of genic molecular markers (GMMs). They are generated from untranslated regions (UTR) of the EST. Zhang et al. (2007) have suggested that UTRs in microsatellite sequences are more polymorphic than the coding regions. However, the ability of UTRs to show any association with the phenotype is compromised because these regions never translate and markers based on UTRs are not functional. Another reason that low marker-trait associations are compromised would be if primers have tri-nucleotide motifs. Polymorphism in microsatellite is caused by slippage, which occurs in the multiples of repeat unit. Therefore, for a tri-nucleotide, the slippage causes a difference in the nucleotides by a multiple of 3. Since a codon is also a

tri-nucleotide, there is no possibility of frame shift mutation. Alteration in protein, therefore, depends only on the size of the slippage in this situation. The other possibility of not finding association could be due to the fact that the phenotyping methods used in this study did not represent all the genetic variations related to the drought response pathways.

Fifty-six polymorphic amplification products were obtained from EST-SSR primers and 11 polymorphisms were displayed by TRAPs. Twenty-two SSR markers and all polymorphic TRAP markers showed association with at least one phenotype in either year, which confirms that the physiological parameters used in this study for phenotyping are comprehensive and correctly reflect the drought pathway. It could also be possible that many of these markers are showing association with the traits due to the large linkage disequilibrium in sugarcane (Janoo et al. 1994; Raboin et al. 2008).

Several marker loci were found associated with more than one physiological parameter consistently for two years (Tables 1, 2, 3, 4). This indicates the possibility of a crosstalk between the signaling pathways of these physiological parameters. The individual effect of these markers appears negligible, but their involvement in multiple physiological parameters involving different sets of genes suggests that they may have a bigger interaction effect. Therefore, allelic variation among these markers could have a major impact on crop production under water limiting conditions. This makes them an ideal candidate for marker assisted selection of abiotic stress tolerance. Drought is a major abiotic stress which affects several metabolic pathways. Drought tolerance is therefore an outcome of up-regulation and down-regulation of different sets of genes, which varies with the environment. The significant association of a marker locus with many physiological parameters (metabolic pathways) makes it effective for selection in different environments. It has been demonstrated that multiple signaling pathways can be activated during exposure to stress, such as drought, cold and salinity, leading to similar responses to different triggers. For example, drought, low temperature, and high salinity, three common abiotic stresses, were all reported to cause an accumulation of compatible solutes and antioxidants (Hasegawa et al. 2000). An overlap in gene

expression between environmental stresses has also been shown in rice (Rabbani et al. 2003). A gene marker which belongs to one of these common metabolic pathways and is triggered by all three stresses can be used as a common marker for selection of tolerance against abiotic stress.

Marker loci showing significant association with multiple physiological parameters such as SSR 9 are probably associated with important genes that are key regulators of many pathways. This finding is in accordance with previous literature (Fujita et al. 2006) suggesting the cross-talk in the stress signaling pathways. SSR 9a is showing association with chlorophyll content as well as relative water content. The EST sequence for SSR 9 is homologous to the protein phosphatase 2 C (PP2C). PP2C is an important component of abscisic-acid (ABA) signaling (Rodriguez 1998; Schweighofer et al. 2004). According to Sheen (1998) AtPP2CA, the Arabidopsis PP2C acts as a negative regulator by blocking ABA signal transduction when transiently expressed in maize mesophyll protoplasts. Down-regulation of AtPP2CA using antisense gene mutant accelerated plant development and led to freezing tolerance (Tahtiharju and Palva 2001). In the current study as well, PP2C is showing a negative correlation with the phenotypes. PP2C has also been reported as a negative regulator in ABA-mediated stomatal closure (Allen et al. 1999). Even though drought responsive stomata conductance is one of the major factors controlling LT, the locus SSR9a did not show any association with LT. LT is a complex trait controlled by many genes. Therefore, it is possible that the main effect of PP2C alone is not noticeable. This could also be possible due to the functional redundancy of the PP2C gene family in the stomatal conductance pathway.

Stomata closure works both ways in terms of drought tolerance. Plants, which fail to close stomata under drought, are susceptible as they desiccate due to excessive transpirational loss. Additionally, plants which are sensitive to drought and close stomata early, do not survive due to heat and failure to transpire. This explains how QTLs interact with environment and why the effects of the markers in this study are not similar for the two years. No marker-trait association was observed with LT which is in

corroboration with Silva et al. (2007), where effect due to genotype was not significant for LT.

SSR 80 is homologous to a cystein protease, which is a component of the proteinase inhibitor complex in plants. Cystein proteases are widely assumed be involved in defense. Their possible functions include the regulation of endogenous protein turn-over (Arai et al. 2002; Corre-Menguy et al. 2002), programmed cell death and plant defense (Zhao et al. 1996; Kuroda et al. 1996; Delledonne et al. 2001).

In this study, SSR 230, which is homologous to patatin, showed significant but negative association with CC, LT, and LRWC. Proteins with sequence homology with patatin were cloned from cucumber seedlings, where a Phospholipase A activity was demonstrated (May et al. 1998) and from *Hevea brasiliensis* latex (Sowka et al. 1998), where esterase activity was shown. Drought induces a decrease in polar lipid content of the plasma membrane (Monteiro et al. 1990). In *Vigna ungiculata* leaves, phospholipid amounts were shown to decrease when watering was suppressed. This is mainly due to the action of phospholipase D (El Maarouf et al. 1999) which is more pronounced in drought-sensitive cultivars. Regarding the major lipid components of chloroplast membranes, the degradative process results from the action of a lipolytic acylhydrolase (Sahsah et al. 1994), an enzyme that removes fatty acids from both *sn* positions.

In this study several TRAP markers, representing DREBs, showed association with leaf relative water content. The DREB transcription factors are associated with the ABA-independent signal transduction pathway. Being dehydration-responsive, such factors are important in abiotic stress pathways. This corroborates these findings that the TRAP markers developed from DREB ESTs are robust. Although TRAPs were considered as less effective than SSRs in detecting polymorphism on the D genome of wheat (Liu et al. 2005), they are generally more polymorphic and of high utility in detect ing the polymorphisms in those ESTs, which does not contain microsatellite sequences.

Interestingly, the EST-SSR 9, SSR 80 and SSR 30 were found significantly associated across the environments. These associations were also validated through permutation tests and simple linear regressions. These results indicate that these markers

are associated with prominent genes in the stress pathways. Association of these markers with more than one trait indicates that they could possibly be involved in cross-talk of stress signaling pathways.

Using these candidate gene based markers for planning crosses would be another practical application of these markers. Screening sugarcane genotypes with these markers may help in avoiding selection of parents susceptible to drought. In fact such markers can facilitate the selection of parents with unique alleles at different loci for hybridization. Hybridization of such genotypes can be a good strategy for pyramiding of drought tolerant alleles. Selection of parents using candidate gene-based markers is already a practice in South Africa (Jorge Da Silva, personal comment). Application of EST-SSR developed in this study for planning crosses has been explained in the next chapter.

Overall, the results found in this study are promising and indicate a useful role of genic markers in the breeding drought tolerant genotypes in sugarcane. However, despite several measures taken to prevent the effect of population structure, there is still a possibility that residual structure remained in the groups identified and could be causing spurious associations between marker and trait. A family-based experimental design would be useful in validating the marker-trait associations which are obtained in this study.

CHAPTER IV

ASSESSMENT OF GENETIC DIVERSITY USING DROUGHT-RELATED GENIC MARKERS

Introduction

Genetic variation is the basis for improvement in any breeding program. Selection in the progeny will be inefficient if genetic variation is not enough in the parents. Molecular markers are good tools for exploring genetic diversity. However, if interest is in a specific trait, the genomic region controlling that trait needs to be explored.

Modern sugarcane varieties are mainly derived from inter-specific crosses between the noble cane *Saccharum officinarum,* which is a high sugar producer and wild species, and *S. spontaneum,* known for its resistance to pathogens and other stresses (Pinto et al 2006). Successive back-crosses have been carried out under the process of nobilization to recover the traits of *S. officinarum* (Bremer 1961). This constant selection of agronomic traits has created, over time, a narrow genetic base causing a plateau in productivity and blocking further gain in yield. Currently, this represents a major concern for sugarcane breeders (Arceneaux 1968; Tai and Miller 1978; Roach and Daniels 1987; Lu et al. 1994; Deren 1995). Exotic germplasm is often used to overcome this plateau in productivity, but this brings many undesired traits along with the desired ones.

The choice of parents to cross is, therefore, a crucial step in a sugarcane improvement program and it requires knowledge and understanding of the genetic diversity of the available germplasm. Several methods have been used to investigate the genetic variation of this crop. Traditional methods rely mainly on the use of pedigree records and phenotypic traits (Skinner et al. 1987; Stevenson 1965). However, phenotyping of quantitative traits is highly influenced by environmental factors and hence, does not reflect the real diversity of *Saccharum* spp. germplasm. The coefficient of parentage (f) is an important method to estimate the genetic diversity based on

pedigrees (Kempthorne 1957). It indirectly measures the genetic diversity among cultivars by estimating, from a pedigree record, the probability that alleles in a locus are identical by descent. However, assumptions regarding the relatedness of ancestors, selection pressure, and genetic drift are generally not met (Lima et al. 2002). The assumption that all genotypes are unrelated may be misleading in cultivated sugarcane where only a handful of clones have been used in the original synthesis.

Molecular markers offer a direct insight at the DNA level. Several molecular marker systems, such as RFLP (Lu et al. 1994), RAPD (Burner et al. 1997), and AFLP (Besse et al. 1998) have been routinely used to assess genetic diversity in sugarcane. These techniques allow amplification of random portions of the genome resulting in competent estimates of genetic diversity. However, these random markers are usually 10-20 cM away from the trait, which leaves a high possibility of recombination between marker and the trait locus. Therefore, the use of markers which co-segregate with the trait of interest may be of more interest to the breeder.

Comparative studies between genomic simple sequence repeats (gSSRs) and expressed sequence tags (EST-SSRs), which are also called genic SSRs reveals that the later displays lower polymorphisms and are not as efficient as gSSRs for distinguishing closely related genotypes (Gupta and Varshney 2000). Furthermore, the development of genic SSRs is restricted to those species which have sufficient sequence data (for ESTs or genes) available. Nevertheless, EST-SSR markers are highly transferable in the related species as they target conserved gene sequences. Genic SSR and genomic SSR markers tend to be complementary for genome mapping. Genomic SSRs display higher polymorphisms and are therefore superior for fingerprinting or varietal identification studies. Genic SSRs shows less polymorphism and are useful for exploring functional genetic diversity (Varshney et al. 2005).

To better understand the genetic diversity in sugarcane cultivars that are useful for cultivar identification and mapping, SSR primers have been designed from ESTs derived from a sugarcane public database (SUCEST-http://sucest.lbi.dcc.unicamp.br/en/). This study includes the application of these markers

to evaluate genetic diversity among 80 sugarcane cultivars from Texas, Louisiana and Florida.

The main limitation of this approach is to apply a large number of markers on the whole population, only a fraction of which are polymorphic. To avoid the application of non-polymorphic markers on large populations, a strategy, which is based on pooling DNA of different genotypes, was developed. Bulking has been efficiently used in other SSR-based studies to assess gene frequencies (Collins et al. 2000). In the present study, we designed bulks by pooling the DNA of extreme drought tolerant and drought susceptible sugarcane genotypes. We previously classified these genotypes according to their tolerance or susceptibility to drought, based on phenotypic evaluation of four physiological parameters: chlorophyll content (CC), chlorophyll fluorescence (CF), leaf temperature (LT) and leaf relative water content (LRWC), used by Silva et al. (2007).

The development of microsatellite markers derived from EST opens new opportunities for genetic investigations at a functional level. However, drought tolerance is a polygenic trait that involves a large number of genes (Yan et al. 2009). Therefore, several candidate genes can be studied. We applied on our population, a set of 55 stress-related EST-SSR primers based on stress related ESTs. Thirteen EST-SSRs showing scorable polymorphisms were selected for estimation of genetic similarity in the current study (Appendix E). Our findings indicate that there is enough variation in this population at drought-related loci. This study also confirms the efficiency of EST-SSR markers for exploring locus-specific genetic diversity in sugarcane.

Materials and Methods

Plant materials

Eighty genotypes of sugarcane were used to evaluate the polymorphism generated by genic markers. These included elite clones and cultivars collected by AgriLife Research-Weslaco from Texas, Louisiana and Florida, USA (Appendix A).

DNA extraction and PCR

Genomic DNA was extracted from 50 mg fresh leaf tissue using the CTAB method (Doyle and Doyle 1990). PCR was performed using a PCTC-100 thermocycler (MJ Research, Inc. Waltham, MA, USA). PCR reaction and PCR conditions has been described under the "Materials and Methods" section in Chapter III.

Genic markers

The procedure of developing EST-SSR primers, which were used in this study, has been described under the "Materials and Methods" section in Chapter III.

Polymorphism analysis

Although SSRs are considered co-dominant markers, but they were treated as dominant markers in this study. The genotyping of sugarcane clones was done based on presence (1) or absence (0) of polymorphic bands. The amplified fragments produced by the EST-SSRs were considered as alleles of a single locus. The polymorphic information content for each locus was calculated according to Cordeiro et al. (2003). The discrimination power was calculated based on Tessier et al. (1999) and used as a measure of SSR loci efficiency for variety identification.

Genetic similarity estimate

Alleles were scored in a dominant manner as presence (1)-absence (0) to construct a matrix, using only polymorphic fragments. This matrix was used to estimate the genetic similarity between all the clones evaluated, adopting the Jaccard (S_j) coefficient (Sneath et al. 1973): $S_j = A/ (A+B+C)$, where A is the number of bands common to both genotypes being compared, B is the number of bands unique to the first genotype, and C is the number of bands unique to the second genotype. The similarity between pairs of genotypes was calculated as the number of alleles in common for the pair, divided by the total number of bands scored for the pair. The genetic relationships among the clones were visualized by a dendrogram based on the unweighted pair group method with

arithmetic averages, UPGMA (Meyer et al. 2004). The cophenetic coefficient (r_{coph}) was calculated to measure the correspondence between the genetic similarity matrix and the cluster analyses (dendrogram). The above analyses were performed with NTSYS-PC software, version 2.2 (Exter Software, NY, USA; Rohlf 1993). A bootstrap method with 1000 replicates was performed to verify if the number of polymorphic markers used for genetic similarity (GS) estimation was enough to support precise estimates among the sugarcane clones evaluated by the program (Lima et al. 2002). This analysis was performed using the Dboot software (Coelho 2000).

Results

Polymorphisms generated by the drought-related EST-SSRs

The 55 EST-SSRs were initially tested on bulk samples of drought tolerant and drought susceptible genotypes to verify the quality of product amplification. These bulk samples were created by pooling the DNA of ten (five at each side) genotypes having extreme values of chlorophyll content (CC), chlorophyll fluorescence (CF), leaf temperature (LT) and leaf relative water content (LRWC). These four physiological parameters have been successfully used for quick screening of drought tolerant genotypes in a large sugarcane population (Silva et al. 2007). Thirteen EST-SSRs, which were polymorphic for bulk DNA, were selected for further study (Appendix E). All of these EST-SSRs showed homology to genes involved in drought-related pathways (Appendix D). The 13 EST-SSRs produced 56 polymorphic alleles. The allele number ranged from 1 (SSR 4, SSE 924) to 9 (SCA 48) with a mean of 4.3 alleles per EST-SSR. Most of the markers showed high values of polymorphic information content (PIC) with an average of 0.54. The greatest PIC value was observed for SCA48 (0.84), whereas the lowest one (0.00) was for EST-SSR4 and EST-SSR924.

Establishment of genetic similarity among sugarcane genotypes based on drought-related EST-SSRs

Genetic similarities based on Jaccard's coefficient were obtained for all possible 3160 pairwise comparisons calculated with 56 polymorphic markers amplified by 13 EST-SSRs. These 3160 pair-wise comparisons gave an average GS value of 0.40. Genetic similarity values varied from 0.03 for genotypes (HOCP00-960 and TCP02-4621) to 0.77 for genotypes (H96-133 and TCP02-4618).

Clustering of sugarcane genotypes by EST-SSR relatedness

A dendrogram was constructed to represent the genetic relationship among the sugarcane clones evaluated (Fig. 3). The co-phenetic value was noted to be high (r = 0.71), indicating a good fit with genetic similarity values. The dendrogram showed that the panel of germplasm selected was diverse at these drought-related loci, and that the genotypes are not distributed based on their pedigree or demographic regions.

Discussion

In the present work, polymorphic markers have been obtained with a subset of 56 EST-SSRs, based on 13 candidate genes involved in drought pathways. The parameters taken into consideration were the number of alleles, polymorphic information content and ability of EST-SSRs to establish genetic relationships among sugarcane clones.

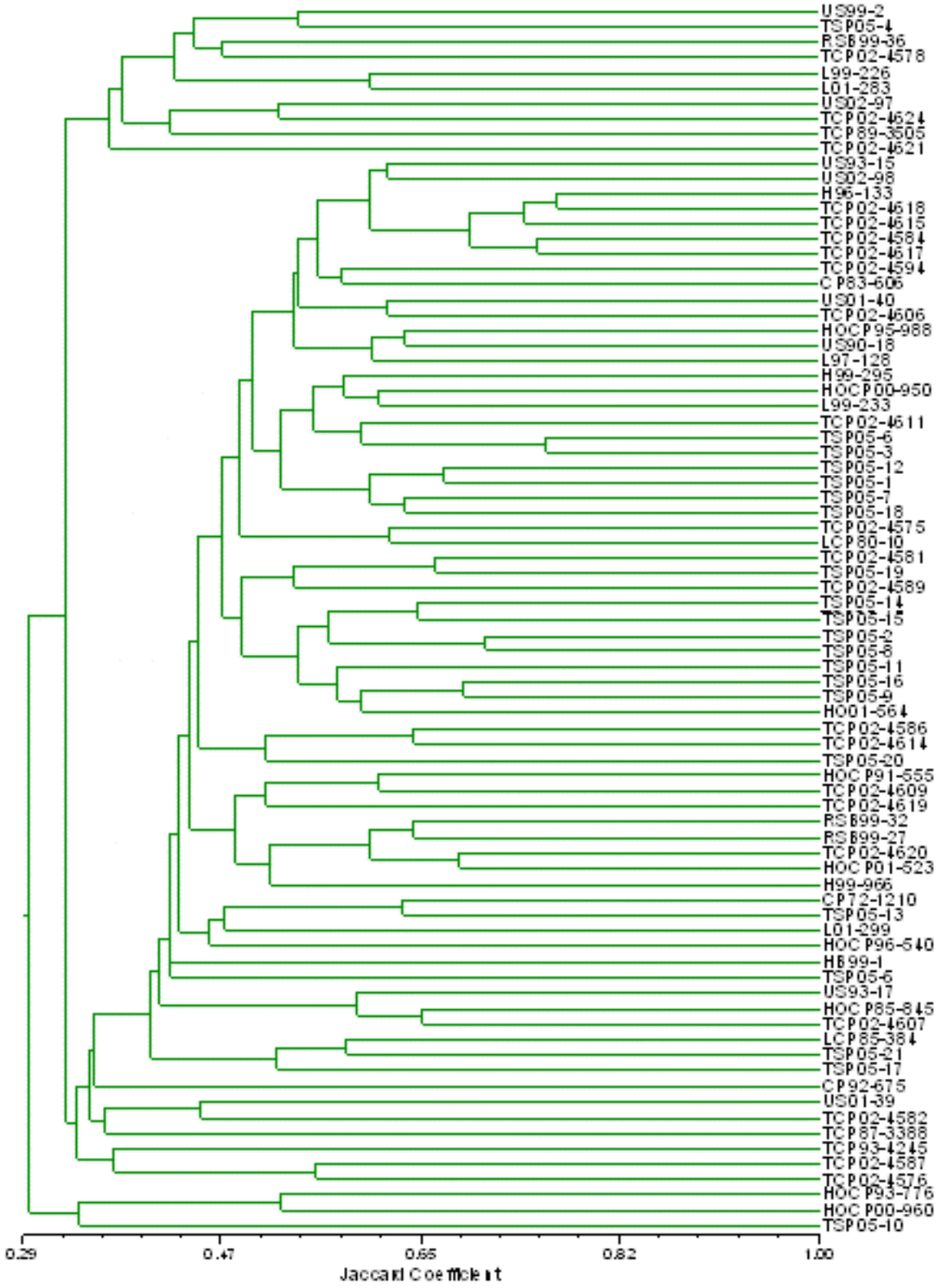

Fig. 3 Dendrogram based on Jaccard coefficient and UPGMA clustering method and derived from EST-SSRs data

Results from the EST-SSR based analyses of 80 modern sugarcane cultivars confirm that modern sugarcane exhibits a high degree of DNA polymorphism when assessed with markers obtained from drought-related genes. The dendrogram clearly discriminates between the genotypes (Figure 3). Even the genotypes that have shared ancestry are seen far from each other at many places, which indicate that the diversity analysis is not influenced by pedigree. This is further supported by the fact that all TSPs (from Sao Paulo) or CPs (from Canal Point) did not form a group in cluster analysis, which is expected based on their common origin. If the genotypes are distributed in these groups as a result of cluster analyses, it would indicate the presence of marker group interaction and absence of significant marker effect within the groups which could be due to population structure effect. Absence of this grouping, however, indicates the presence of within-groups variance. Therefore, this supports the hypothesis that marker-trait association is independent of population structure (Wei et al 2006). Although the effect of population structure cannot be completely determined by the dendrogram, which is a two-dimensional figure, it gives an indication that this population could be suitable for association mapping studies.

Molecular markers derived from gene sequences commonly displays low level of polymorphism, which can limit their widespread use in genetic analysis (Varshney et al. 2005). However, inspite of the narrow genetic base of sugarcane (*Saccharum* spp), 13 primers were sufficient to distinguish the 80 genotypes in this study. This result demonstrates that EST-SSRs are useful for generating reliable genetic parameter estimates for assessment of locus-specific variability.

The genetic similarity (GS) estimated in this study was low compared to previous studies in sugarcane because the EST-SSRs were locus-specific and, thus, they amplified the region related to stress-responsive genes. Therefore, over 90% of the fragments amplified by these markers were polymorphic. The high number of polymorphic alleles produced by these primers could be due to lack of selection for stress tolerance in these genotypes.

Estimate of GS is used as a tool to assist sugarcane breeders with selecting the most divergent parents to maximize heterosis and transgressive segregation in the progeny population. The genetic and evolutionary structure of populations can be studied using a technique known as numerical taxonomy. Similarity coefficients are used for numerical taxonomy and those using binary data, such as the one utilized in this study. These are referred to as association coefficients (Sneath and Sokal 1973).

Molecular markers are considered as genetic traits that are selectively neutral, free of epistatic interactions and unaffected by environment (Tanksley et al. 1989). Microsatellite (SSR) markers that are based on candidate genes could be a valuable tool to estimate the trait specific variation. Determination of variation in the regions controlling a desired trait may improve the precision in selecting the parents for a breeding program. As such, SSRs are particularly well suited for taxonomic resemblance, one of the areas of numerical taxonomy. However, when using markers for taxonomic resemblance studies, two important mathematical assumptions are made by choosing markers sufficiently spaced on the genetic map for the species: 1) Independence among markers and 2) Equal weight of markers in creating taxa. Because SSRs utilized in this study have not, as of date, been placed in any sugarcane genetic map, one has to bear in mind, when examining the dendrogram being presented (Figure 3), that the above-mentioned mathematical assumptions cannot be made.

Genotypes in the dendrogram were not completely distributed based on their phenotypic response to drought-related loci (Figure 3). An investigation was carried out to see if already known drought tolerant cultivars are located closely in the dendrogram. TCP02-4620 and HOCP01-523 that are known drought tolerant genotypes, were found clustered very closely in the dendrogram (Figure 3) but such examples were exceptions. In most cases, tolerant and susceptible genotypes were found intermingled together. It is important to note that our study does not focus only on commercial cultivars. It also includes several genotypes, which have not undergone enough selections for alleles with negative effects to be removed. Genetic diversity is estimated by physical separation of DNA fragments based on their size without taking into account how they interact with

each other and with the environment. Interpretation of DNA fragments could reflect the alleles of either positive or negative impact on the phenotype. Total genetic diversity between two genotypes at a number of loci may therefore have a weak correlation with the phenotype, unless alleles are in the correct combination. In addition to this, there could be several general factors causing disparity between cluster analysis and phenotypic evaluation such as strong non-additive effects and/or involvement of genetic factors other than these 13 loci. Nevertheless, this diversity in genic regions is useful in making a choice of the parents to cross and it encourages us to conduct association analyses on an individual marker basis and determine the models, which can later be used for marker assisted selection for drought tolerance.

CHAPTER V

CONCLUSIONS

This investigation has produced important information for marker assisted selection of drought tolerance in sugarcane and other related species. Molecular markers developed in this study may also be applicable in other abiotic stress. Fifty-six polymorphisms produced by SSR markers and 14 polymorphisms produced by TRAP markers can be used for fingerprinting in sugarcane genotypes. Population structure was determined in this sugarcane population and its effect was removed. Therefore this population is good for other association studies. This study confirms that complex traits like drought tolerance can be dissected in sugarcane through association study.

In year 2008, the period of drought treatment was short compared to 2006. However, the marker-trait associations detected in this study were much higher in 2006. This indicates that EST based markers are precise and the associations detected are related to the trait expression. Not many associations are lost in 2008 in the case of TRAP markers, which is probably because all TRAP markers were based on DREB homologous sequences. DREB are transcription factors, therefore they are most likely early response elements. All markers may be genuine but markers showing association in 2008 may be associated with early responding genes compared to those which did not express in 2008.

Quantification of genetic variability of genotypes from a Texas sugarcane breeding program at drought specific loci was done by GS values. Knowing the difficulties in obtaining superior genotypes in sugarcane breeding programs, due to narrow genetic base available, the selection of parents with similarity values smaller than average genetic similarity (0.40) is recommended.

REFERENCES

Agarwal PK, Agarwal P, Reddy MK, Sopory SK (2006) Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. Plant Cell Rep 25:1263-1274

Aguilera C, Stirling CM, Long SP (1999) Genotypes variation within *Zea mays* for susceptbility to and rate of recovery from chill-induced photoinhibition of photosynthesis. Physiol Plant 106:429-436

Allen GJ, Kuchitsu K, Chu SP, Murata Y, Schroeder JI (1999) *Arabidopsis* abi1-1 and abi2-1 phosphatase mutations reduce abscisic acid-induced cytoplasmic calcium rises in guard cells. The Plant Cell 11:1785-1798

Altinkut A, Kazan K, Ipekci Z, Gozukirmizi N (2001) Tolerance to paraquat is correlated with the traits associated with water stress tolerance in segregating F2 populations of barley and wheat. Euphytica 121:81-86

Arai S, Matsumoto I, Emori Y, Abe K (2002) Plant seed cystatins and their target enzymes of endogenous and exogenous origin. J Agric Food Chem 50:6612-6617

Arceneaux G (1968) Cultivated sugar canes of the world and their botanical derivation. Proc Int Sugar Cane Technol 12:844-854

Besse P, Taylor G, Carroll B, Berding N, Burner D, McIntyre CL (1998) Assessing genetic diversity in a sugarcane germplasm collection using an automated AFLP analysis. Genetica 104:143-153

Blum A (2005) Drought resistance, water use efficiency, and yield potential: are they compatible dissonant, or mutually exclusive? Aust J Agric Res 56:1159-1168

Bremer G (1961) Problems in breeding and cytology of sugarcane. Euphytica 10:59-78

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165-1177

Burner DM. Pan YB, Webster RD (1997) Genetic diversity of North American and Old World *Saccharum* assessed by RAPD analysis, Genet Resour Crop Evol 44:235-240

Butterfield MK, Rutherford RS, Carson DL and Huckett BI (2004) Application of gene discovery to varietal improvement in sugarcane. S Afr J Bot 70:167

Campos H, Cooper M, Habben JH, Edmeades GO, Schussler JR (2004) Improving drought tolerance in maize: a view from industry. Field Crop Res 90:19-34

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963-971

Coelho SG (2000) Software DBOOT – Avaliação dos erros associados a estimativas de distâncias/similaridades genéticas através do procedimento de bootstrap com número variável de marcadores, v. 1.1. Goiânia, Brasil, Universidade Federal de Goiás cold acclimation in *Arabidopsis thaliana*. Plant J 26(4):461-70

Collins HE, Gamba G, Pontieri S, Farina A, Gaudio L (2000) Isolation and characterization of microsatellite total allele content differences between DNA pools. Hum Genet 106(2):218-226

Colom MR, Vazzana C (2003) Photosynthesis and PSII functionality of drought-resistant and drought sensitive weeping lovegrass plants. Environ Exp Bot 49:135-144

Cooper M, Podlich DW, Luo L (2007) Modeling QTL effects and MAS in plant breeding. In: Varshney RK, Tuberosa R (eds) Genomic Assisted Crop Improvement Genomics Approaches and Platforms. Springer, New York, pp 57–95

Cordeiro GM, Pan YB, Henry RH (2003) Sugarcane microsatellites for assessment of genetic diversity in sugarcane germplasm. Plant Sci 165:181-189

Corre-Menguy F, Cejudo FJ, Mazubert C, Vidal J, Lelandais-Briere C, Torres G, Rode A, Hartmann C (2002) Characterization of the expression of a wheat cystatin gene during cariopsis development. Plant Mol Biol 50:687-698

Daugrois JH, Grivet L, Roques D, Hoarau JY, Lombard H, Glaszmann JC and D'Hont, A (1996) A putative major gene for rust resistance linked with a RFLP marker in sugarcane cultivar 'R570'. Theor Appl Genet 92:1059-1064

Delledonne M, Allegro G, Belenghi B, Balestrazzi A, Picco F, Levine A, Zelasco S, Calligari P, Confalonieri M (2001) Transformation of white poplar (*Populus alba* L.) with a novel Arabidopsis thaliana cysteine proteinase inhibitor and analysis of insect pest resistance. Mol Breed 7:35-42

Deren CW (1995) Genetic base of U.S. mainland sugar cane. Crop Sci 35: 1195-1199

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13-15

Earl HJ, Tollenaar M (1999) Using chlorophyll fluorometry to compare photosynthetic performance of commercial maize (*Zea mays* L.) hybrids in the field. Field Crops Res 61:201-210

El Maarouf H, Zuily-Fodil Y, Gareil M, d'Arcy-Lameta A, Pham-Thi AT (1999) Enzymatic activity and gene expression under water stress of phosphatase D in two cultivars of *Vigna ungiculata* L. Walp. Differing in drought tolerance. Plant Mol Biol 39:1257-1265

Ersoz ES, Yu J, Buckler ES (2007) Applications of linkage disequillibrium and association mapping in crop plants. In: Varshney RK, Tuberosa R (eds) Genomics-Assisted Crop Improvement: Vol 1 Genomics Approaches and Platforms. Springer, The Netherlands, pp 97-119

Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-Schinozaki K and Shinozaki K (2006). Crosstalk between abiotic and biotic stress responses: a current view from the point of convergence in the stress signaling networks. Curr Opin Plant Biol 9:436-442.

Gascho GJ, Shih SF (1983) Sugarcane. In: Teare ID and M.M. Peet (eds) Crop Water Relations. John Wiley & Sons, New York, pp 445-479

Gimenez C, Mitchell VJ, Lawlor DW (1992) Regulation of photosynthetic rate of two sunflower hybrids under water stress. Plant Physiol 96:635-643

Grivet L, D'Hont A, Roques D, Feldmann P, Lanaud C, Glaszmann JC(1996) RFLP mapping in cultivated sugarcane (*Saccharum* spp.): genome organization in a highly polyploidy and aneuploid interspecific hybrid. Genetics 142:987-1000

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetics and plant breeding with emphasis on bread wheat. Euphytica 113:163-185

Hackauf B, Wehling P (2002) Identification of microsatellite polymorphisms in an expressed portion of rye genome. Plant Breed 121:17-25

Hasegawa M, Bressan R, Pardo JM (2000) The dawn of plant tolerance genetics. Trends Plant Sci 5:317-9

Hogarth D (1987) Genetics of sugarcane. In: Heinz DJ (ed) Sugar Improvement through Breeding. Elsevier Press, Amsterdam, pp 255-271

Hu J, Vick BA (2003) Target Region Amplification polymorphism: a novel marker technique for plant genotyping. Plant Mol Biol Rep 21:289-294

Ingram J, Bartels D (1996) The molecular basis of dehydration tolerance in plants. Annu Rev Plant Physiol Plant Mol Biol 47:377–403

Jamaux I, Steinmertz A, Belhassen E (1997) Looking for molecular and physiological markers of osmotic adjustment in sunflower. New Phytol 137:117-127

Jannink JL, Bink M, Jansen RC (2001) Using complex plant pedigrees to map valuable genes. Trends Plant Sci 6:337-342

Jannink JL, Walsh B (2002) Association mapping in plant population. In: Kang MS (ed) Quantitative Genetics, Genomics and Plant Breeding. CABI, Wallingford, UK, pp 59-68

Jannoo N, Grivet L, Dookun A, D'Hont A, Glaszmann JC (1999) Linkage disequilibrium among modern sugarcane cultivars. Theor Appl Genet 99:1053-1060

Kempthorne O (1957) An Introduction to Genetic Statistics. John Wiley and Sons, New York

Kuroda M, Ishimoto M, Suzuki K, Kondo H, Abe K, Kimura K, Arai S (1996) Oryzacystatins exhibit growth-inhibitory and lethal effects on different species of bean insect pests, *Callosobruchus chinensis* (Coleoptera) and *Riptortus calavatus* (Hemiptera). Biosci Biotechnol Biochem 60:209-212

Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL, Quackenbush J (2000) An optimized protocol for analysis of EST sequences. Nucleic Acids Res 28:3657-3665

Lima MLA, Garcia AAF, Oliveira KM, Matsuoka S, Arizono H, de Souza CL Jr, de Souza AP (2002) Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). Theor Appl Genet 104:30-38

Liu PY, Zhu J and Yan Lu (2003) A method for marker-assisted selection based on QTLs with epistatic effect. Genetica 119:75-86

Liu ZH, Anderson JA, Hu J, Friesen TL, Ramussen JB, Faris JD (2005) A wheat intervarietal genetic linkage map based on microsatellite and target region amplified polymorphism markers and its utility for detecting quantitative trait loci. Theor Appl Genet 111:782-794

Lu YH, D'Hont A, Paulet F, Grivet L, Arnaud M, Glaszmann JC (1994) Molecular diversity and genome structure in modern sugar cane varieties. Euphytica 78:217-226

Matin MA, Brown JH, Ferguson H (1989) Leaf water potential, relative water content, and diffusive resistance as screening techniques for drought resistance in barley. Agron J 81:100-105

Maxwell K, Johnson GN (2000) Chlorophyll fluorescence-a practical guide. J Exp Bot 51:659-668

Markwell J, Osterman JC, Mitchell JL (1995) Calibration of the Minolta SPAD-502 leaf chlorophyll meter. Photosynth Res 46:467-472

May S, Preisig-Müller R, Hőne M, Grau P, Kindl, H (1998) Biochim Biophys Acta 93: 267-276

Meyer AS, Garcia AAF, Souza AP (2004) Comparison of similarity coefficients used for cluster analysis with dominant markers in maize (*Zea mays*). Genet Mol Biol 27:83-91

Monteiro de Paula F, Pham-Thi AT, Vieira da Silva J, Justin AM, Demandre C, Mazliak P (1990) Effects of water stress on the molecular speciesvcomposition of polar lipids from *Vigna ungiculata* L. leaves. Plant Sci 66:185-193

Moore PH (1995) Temporal and spatial regulation of sucrose accumulation in the sugarcane stem. Aust J Plant Physiol 22:661-679

Nordborg M and Tavare S (2002) Linkage disequilibrium: what history has to tell us. Trends Genet 18:83-90

Pennisi E (2008) The blue revolution drop by drop, gene by gene. Science 320:171-173

Pinto LR, Oliveira KM, Marconi T, Garcia AAF, Ulian EC, de Souza AP (2006) Characterization of novel sugarcane expressed sequence tag microsatellites and their comparison with genomic SSRs. Plant Breed 125:378-384

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215-221

Price S (1965) Interspecifc hybridization in sugarcane breeding. Proc Intl Sugar Cane Technol 12:1021-1026

Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 65:220-228

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959

Qing ZM, Jing LG, Kai CR (2001) Photosynthesis characteristics in eleven cultivars of sugarcane and their responses to water stress during the elongation stage. Proc Intl Soc Sugar Cane Technol 24:642-643

Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, Ito Y, Yoshiwara K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Monitoring expression profiles of rice genes under cold, drought and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. Plant Physiol 133(4):1755-67

Raboin LM, Pauquet J, Butterfield M, D'Hont A, Glaszmann JC (2008) Analysis of genome-wide linkage disequilibrium in the highly polyploidy sugarcane. Theor Appl Genet 116:701-714

Ramesh P (2000) Effect of different levels of drought during the formative phase on growth parameters and its relationship with dry matter accumulation in sugarcane. J Agron Crop Sci 185:83-89

Ribaut JM, Hoisington D, Banziger M, Setter T, Edmeades G (2004) Genetic dissection of drought tolerance in maize: a case study. In JM Ribaut (ed) Physiology and Biotechnology Integration for Plant Breeding, Marcel Dekker, New York, pp 571-609

Risch N and Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516-1517

Roach, B.T. 1972. Nobilization of sugarcane. Proc Int Sugar Cane Technol 14:206–216

Roach BT, Daniels J (1987) A review of the origin and improvement of sugar cane. Copersucar Int Sugar Cane Breeding Workshop, pp 1-31, Copersucar, São Paulo

Rodriguez PL, Benning G, Grill E (1998) ABI2, a second protein phosphatase 2C involved in abscisic acid signal transduction in *Arabidopsis*. FEBS lett 421:185-190

Rohlf FJ (1993) NTSYS-PC numerical taxonomy and multivariate analysis system Version 2.2j. Exeter Publ, Setauket, New York

Rong-hua L, Pei-guo G, Baum M, Grando S, Ceccarelli S (2006) Evaluation of chlorophyll content and fluorescence parameters as indicators of drought tolerance in barley. Agri Sci China 5:751-757

Sahsah Y, Campos P, Gareil M, Zuily-Fodil Y, Pham-Thi AT (1998) Enzymatic degradation of polar lipids in *Vigua ungiculata* leaves and influence of drought stress. Physiol Plant 104: 577-586

Schweighofer A, Hirt H, Meskiene I (2004) Plant PP2C phosphatases: emerging function in stress signaling. Trends Plant Sci 9(5):236-43

Sheen J (1998) Mutational analysis of a protein phosphatase 2C involved in abscisic acid signal transduction in higher plants. Proc Natl Acad Sci USA 95:975-980

Silva JAG, Sorrells ME, Burnquist WL and Tanksley SD (1993) RFLP linkage map and genome analysis of *Saccharum spontaneum*. Genome 36:782-791

Silva JAG (2001) Preliminary analysis of microsatellite markers derived from sugarcane expressed sequence tags (ESTs). Genet Mol Biol 24:155-159

Silva MA, Jifon JL, Da Silva JAG, Sharma V (2007) Use of physiological parameters as fast tools to screen for drought tolerance in sugarcane. Braz J Plant Physiol 19:193-201

Silva MA, Silva JAG, Enciso J, Sharma V, Jifon J, (2008) Yield components as indicators of drought tolerance of sugarcane. Sci Agric (Piracicaba, Braz) 65:620-627

Singh S, Rao PNG (1987) Varietal differences in growth characteristics in sugarcane. J Agri Sci 108:245-247

Skinner JC, Hogarth DM, Wu KK (1987) Selection methods, criteria, and indices. In: Heinz DJ (ed) Sugarcane Improvement through Breeding. Elsevier, Amsterdam, pp 409-453

Sneath, PHA and Sokal, RR (1973). Numerical Taxonomy, The Priniciples and Practice of Numerical Classification. WH Freeman and Company. San Francisco. 573 p

Sowka S, Wagnar S, Krebitz M, Arija-Mad-Arif S, Yusof F, Kinaciyan T, Brhehler R, Scheiner O, Breiteneder H (1998) cDNA cloning of the 43-kDa latex allergen Hey b 7 with sequence similarity to patatins and iits expression in the yeast *Pichia pastoris*. Eur J Biochem 255: 213-219

Sreenivasan TV, Ahloowalia BS, and Heinz DJ (1987) Cytogenetics. In: Heinz DJ (ed) Sugarcane Improvement through Breeding. Elsevier, Amsterdam, pp 211-253

Sterky F, Lundeberg J (2000) Sequence analyses of genes and genomes. J Biotech 76: 1-31

Stevenson, GC (1965) Genetics and Breeding of Sugarcane. Longmans, London

Tähtiharju S, Palva T (2001) Antisense inhibition of protein phosphatase 2C accelerates cold acclimation in *Arabidopsis thaliana*. Plant J 26:461-70

Tai P, Miller JD (1978) The pedigree of selected Canal Point (CP) varieties of sugar cane. Proc Am Soc Sugar Cane Tech 8:34-39

Tanksley SD, Young ND, Paterson AH, Bonierbalf MW (1989) RFLP mapping in plant breeding: new tools for an old science. Bio Technology 7:257-264

Tardieu F (1996) Drought perception by plants. Do cells of droughted plants experience water stress? Plant Growth Regul 20:93–104

Tardieu F (2003) Virtual plants: modeling as a tool for the genomics of tolerance to water deficit. Trends Plant Sci 8:9-14

Tessier C, David J, This P, Boursiquot JM, Charrier A (1999) Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. Theor Appl Genet 98:171-177

Tollenaar M, Aguilera A (1992) Radiation use efficiency of an old and new maize hybrid. Agron J 84:536-541

Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48-55

Wang Z, Weber JL, Zhong G, Tankesley SD (1994) Survey of plant short tandem DNA repeats. Theor Appl Genet 88:1-6

Wei X, Jackson PA, McIntyre CL, Aitken KS, Croft B (2006) Association between DNA markers and resistance to diseases in sugarcane and effects of population structure. Theor Appl Genet 114:155-164

Wolfe DW, Henderson DW, Hsaio TC, Alvino A (1988) Interactive water and nitrogen effect on senescence of maize: II. Photosynthetic decline and longevity of individual leaves. Agron J 80:865-870

Wolfsberg TG, Landsman D (2001) Expressed sequence tags (ESTs). In: Baxevanis AD and Ouellette BFF (eds) Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. John Wiley and Sons, New York, pp 283–302

Yan J, Setter TL, Warburton M, Buckler ES, Sawkins M, Ribaut JM, Xu Y, Grudloyma P, Gethi J (2009) Development of informative markers through association mapping in maize to improve drought tolerance in cereals. Plant and Animal Genome 17: 329

Zhang Z, Deng Y, Tan J, Hu S, Yu J, Xue Q (2007) A genome-wide microsatellite polymorphism database for the *Indica* & *Japonica* rice. DNA Res 14:37-45

Zhao Y, Botella MA, Subramanian L, Niu X, Nielson SS, Bressan RA, Hasegawa PM (1996) Two wound-inducible soybean cystein proteinases have greater insect digestive proteinase inhibitory activities than a consecutive homolog. Plant Physiol 111: 1299-1306

APPENDIX A

Variety	Female	Male
CP72-1210	CP65-357	CP56-63
CP83-605	CP73-340	CP66-346
CP92-675	CP83-644	CP70-321
H96-133	L 90-178	HOCP 85-845
H99-295	HO 94-856	HOCP 93-750
H99-966	HOCP 92-674	HO 94-861
HB99-1	US 98-03	HOCP 95-951
HO01-564	Green German	LCP 85-384
HOCP00-950	HOCP 93-750	HOCP 92-676
HOCP00-960	US 90-18	US 90-17
HOCP01-523	TCP 86-3374	HOCP 85-845
HOCP85-845	CP 72-370	CP 77-403
HOCP91-555	CP 83-644	LCP 82-094
HOCP93-776	HOCP 85-845	CP 84-742
HOCP95-988	CP 86-941	US 89-12
HOCP96-540	LCP 86-454	LCP 85-384
L01-283	L93-365	LCP85-384
L01-299	L93-365	LCP85-384
L97-128	LCP81-010	LCP85-384
L99-226	CP89-846	LCP81-030
L99-233	CP79-348	HOCP91-552
LCP80-10	CP74-328	CP70-1133
LCP85-384	CP77-310	CP77-407
RSB99-27	US 96-2	US 93-16
RSB99-32	US 93-16	HOCP 85-845
RSB99-36	HOCP 85-845	US 90-18

Variety	Female	Male
TCP02-4575	TCP92-4193	?
TCP02-4576	TCP92-4193	?
TCP02-4578	TCP92-4193	?
TCP02-4581	TCP92-4193	?
TCP02-4582	TCP86-3374	LCP86-454
TCP02-4584	CP92-1666	CP93-1634
TCP02-4586	CP92-1666	CP93-1634
TCP02-4587	CP92-1666	CP93-1634
TCP02-4589	TCP90-4095	LCP86-454
TCP02-4594	TCP91-4138	?
TCP02-4606	TCP89-3498	CP92-1167
TCP02-4607	TCP92-4193	TCP91-3543
TCP02-4609	TCP92-3583	TCP89-3505
TCP02-4611	TCP92-3583	TCP89-3505
TCP02-4614	TCP88-3461	TCP86-3368
TCP02-4615	TCP89-3513	TCP91-4138
TCP02-4617	TCP89-3498	TCP88-3461
TCP02-4618	TCP88-3448	HOCP95-929
TCP02-4619	TCP92-3583	TCP88-3461
TCP02-4620	TCP88-3461	TCP91-4138
TCP02-4621	TCP88-3461	TCP91-4138
TCP02-4624	CP90-1424	CP57-614
TCP87-3388	CP70-321	?
TCP89-3505	CP70-321	CP78-304
TCP93-4245	CP70-321	SP70-1143
TSP05-1	TCP98-4450	?
TSP05-10	TCP94-4342	TCP95-4370
TSP05-11	TCP94-4342	TCP95-4370

Variety	Female	Male
TSP05-12	TCP94-4342	TCP95-4370
TSP05-13	TCP94-4342	TCP95-4370
TSP05-14	TCP94-4438	TCP90-4087
TSP05-15	CP89-2377	CP92-1167
TSP05-16	TCP 98-4450	TCP90-4087
TSP05-17	RB72-454	?
TSP05-18	TCP 98-4450	TCP95-4370
TSP05-19	TCP 98-4450	TCP95-4370
TSP05-2	TCP98-4450	?
TSP05-20	TCP93-4245	TCP88-3461
TSP05-21	TCP93-4245	TCP88-3461
TSP05-3	CP92-1167	CP80-1827
TSP05-4	CP92-1167	CP80-1827
TSP05-5	CP92-1167	CP96-1252
TSP05-6	CP92-1167	CP96-1252
TSP05-7	CP92-1167	?
TSP05-8	TCP97-4424	CP96-1252
TSP05-9	TCP91-4138	CP72-1210
US01-39	HoCP 92-678	US 93-15
US01-40	HoCP 93-775	US 93-15
US02-97	US 93-16	US 90-24
US02-98	HoCP 96-569	US 93-17
US90-18	CP 79-348	CP 83-657
US93-15	CP 85-861	CP 85-834
US93-17	LCP 84-222	CP 85-834
US99-2	US 90-26	HOCP 92-678

*? = polycross

APPENDIX B

Variety	CC[a]	CF[b]	LT[c]	LRWC[d]
CP72-1210	5.05	.004	-3.55	2.63
CP83-605	0.00	.029	-3.48	5.55
CP92-675	0.00	.059	-4.50	20.53
H96-133	0.10	.015	-4.55	10.75
H99-295	0.40	.048	-4.49	12.68
H99-966	0.80	.009	-4.37	3.50
HB99-1	1.30	.009	-3.70	7.99
HO01-564	1.30	.022	-3.94	5.94
HOCP00-950	1.35	.029	-5.03	7.73
HOCP00-960	1.48	.003	-3.89	4.29
HOCP01-523	1.53	.003	-2.47	1.59
HOCP85-845	8.73	.006	-3.43	9.32
HOCP91-555	1.53	.008	-3.63	3.65
HOCP93-776	1.55	.047	-6.26	13.85
HOCP95-988	1.58	.034	-4.51	12.32
HOCP96-540	1.63	.010	-4.10	6.56
L01-283	1.72	.044	-2.98	8.71
L01-299	1.83	.002	-4.91	2.77
L97-128	11.23	.005	-3.33	4.22
L99-226	1.85	.025	-2.87	5.78
L99-233	1.90	.033	-3.88	2.85
LCP80-10	2.13	.038	-2.79	5.01
LCP85-384	2.13	.006	-4.64	7.78
RSB99-27	2.23	.028	-3.64	9.76
RSB99-32	2.33	.007	-4.56	3.88
RSB99-36	2.45	.005	-3.55	7.69
TCP02-4575	2.58	.027	-3.81	13.41
TCP02-4576	2.70	.027	-5.84	8.97
TCP02-4578	2.80	.017	-4.87	3.32
TCP02-4581	3.15	.027	-3.61	7.73
TCP02-4582	3.20	.012	-4.82	18.73
TCP02-4584	3.23	.007	-2.43	3.03
TCP02-4586	1.85	.014	-1.68	11.85
TCP02-4587	3.40	.011	-3.61	0.84
TCP02-4589	3.40	.028	-3.15	4.84
TCP02-4594	3.63	.005	-5.87	7.14
TCP02-4606	3.63	.054	-1.06	12.54
TCP02-4607	3.70	.026	-4.22	7.60
TCP02-4609	3.83	.014	-3.50	4.22
TCP02-4611	4.08	.002	-4.67	4.1

Variety	CC[a]	CF[b]	LT[c]	LRWC[d]
TCP02-4614	4.10	.006	-6.33	12.99
TCP02-4615	4.18	.015	-2.13	3.39
TCP02-4617	4.20	.013	-4.08	5.82
TCP02-4618	4.22	.027	-4.71	6.28
TCP02-4619	4.45	.024	-6.05	9.11
TCP02-4620	4.58	.016	-3.12	3.99
TCP02-4621	4.78	.025	-3.72	5.70
TCP02-4624	4.83	.064	-6.10	12.16
TCP87-3388	4.90	.027	-4.61	5.20
TCP89-3505	5.18	.031	-2.79	17.18
TCP93-4245	3.23	.000	-5.58	3.65
TSP05-1	5.18	.016	-4.32	13.60
TSP05-10	5.28	.020	-1.28	4.57
TSP05-11	5.33	.003	-3.92	7.90
TSP05-12	5.40	.008	-2.08	4.49
TSP05-13	5.55	.005	-0.07	1.82
TSP05-14	5.63	.033	-3.88	14.69
TSP05-15	5.83	.041	-1.70	14.12
TSP05-16	5.98	.021	-1.54	8.23
TSP05-17	6.18	.023	-4.45	12.39
TSP05-18	6.50	.010	-3.80	2.96
TSP05-19	6.60	0.030	-4.88	7.24
TSP05-2	6.63	.065	-4.92	6.60
TSP05-20	6.68	.007	-3.88	3.01
TSP05-21	6.88	.030	-3.17	13.89
TSP05-3	6.97	0.030	-6.13	6.71
TSP05-4	7.13	.040	-3.14	7.77
TSP05-5	7.28	.004	-3.76	10.08
TSP05-6	7.35	.008	-3.76	1.85
TSP05-7	7.40	.006	-1.86	2.28
TSP05-8	7.78	.015	-3.85	2.68
TSP05-9	8.10	.029	-3.64	16.48
US01-39	8.13	.031	-5.83	8.66
US01-40	9.65	.015	-1.82	7.45
US02-97	6.65	.079	-4.50	5.84
US02-98	9.45	.023	-1.18	7.90
US90-18	1.53	.001	-4.63	2.22
US93-15	10.75	.022	-4.78	8.73
US93-17	8.68	.012	-1.33	5.52
US99-2	13.05	.017	-3.59	1.47

a-chlorophyll content, b-chlorophyll fluorescence, c-Leaf temperature, d-leaf relative water content

APPENDIX C

Variety	CC[a]	CF[b]	LT[c]	LRWC[d]
CP72-1210	0.00	0.00	0.00	9.44
CP83-605	2.85	0.009	-2.34	17.64
CP92-675	-	0.016	-0.98	4.84
H96-133	1.20	0.00	-1.57	11.01
H99-295	0.00	0.004	-10.39	12.57
H99-966	0.00	0.00	-0.62	14.55
HB99-1	0.00	0.070	-2.20	14.62
HO01-564	0.00	0.017	0.00	11.71
HOCP00-950	1.55	0.00	-3.80	0.67
HOCP00-960	0.00	0.00	-2.63	0.98
HOCP01-523	4.55	0.00	0.00	4.85
HOCP85-845	0.00	0.00	-0.34	12.38
HOCP91-555	5.95	0.025	-0.64	-
HOCP93-776	7.10	0.016	0.00	-
HOCP95-988	7.35	0.021	-1.82	2.54
HOCP96-540	0.00	0.048	-1.37	22.12
L01-283	5.75	0.00	0.00	10.34
L01-299	-	0.00	-1.08	2.60
L97-128	0.80	0.028	0.00	0.84
L99-226	1.30	0.00	-0.15	1.17
L99-233	0.85	0.00	0.00	1.75
LCP80-10	0.00	0.00	-1.28	7.60
LCP85-384	3.15	0.008	0.00	13.22
RSB99-27	0.00	0.023	-1.54	2.23
RSB99-32	0.00	0.00	-3.23	4.87
RSB99-36	0.00	0.004	-3.87	3.73
TCP02-4575	12.75	0.00	-2.79	19.39
TCP02-4576	0.60	0.011	-4.98	10.62
TCP02-4578	0.30	0.035	-1.23	12.54
TCP02-4581	8	0.00	-0.43	12.90
TCP02-4582	2.35	0.002	-1.18	-
TCP02-4584	0.00	0.00	-1.46	11.66
TCP02-4586	2.70	0.00	-3.04	5.0
TCP02-4587	0.90	0.00	-0.02	10.52
TCP02-4589	1.95	0.00	-0.79	15.43
TCP02-4594	0.05	0.00	0.00	15
TCP02-4606	5.05	0.00	-0.45	20.07
TCP02-4607	11.85	0.00	-0.81	6.50
TCP02-4609	0.00	0.00	0.00	12.01
TCP02-4611	0.00	0.045	-2.45	7.80

Variety	CC[a]	CF[b]	LT[c]	LRWC[d]
TCP02-4614	0.00	0.00	-0.89	22.03
TCP02-4615	0.00	0.00	0.00	5.27
TCP02-4617	0.00	0.00	0.00	6.38
TCP02-4618	7.25	0.00	-1.69	8.72
TCP02-4619	14.75	0.00	-1.18	-
TCP02-4620	0.00	0.00	-1.50	10.21
TCP02-4621	4.60	0.00	-0.74	9.32
TCP02-4624	0.00	0.014	-1.43	12.56
TCP87-3388	0.00	0.00	-1.86	5.73
TCP89-3505	5.60	0.00	-0.03	-
TCP93-4245	1.85	0.00	-7.45	4.55
TSP05-1	0.00	0.002	0.00	7.14
TSP05-10	8.0	0.00	0.00	3.14
TSP05-11	0.00	0.00	-3.98	4.51
TSP05-12	5.40	0.005	-0.74	8.85
TSP05-13	1.10	0.007	0.00	7.11
TSP05-14	-	0.00	-1.51	0.84
TSP05-15	0.00	0.00	0.00	-
TSP05-16	-	0.00	0.00	17.13
TSP05-17	1.10	0.00	0.00	6.09
TSP05-18	0.55	0.00	-1.55	8.89
TSP05-19	3.70	0.00	-0.26	1.21
TSP05-2	0.00	0.00	0.00	3.24
TSP05-20	0.00	0.006	-1.86	6.37
TSP05-21	0.00	0.00	-0.30	2.85
TSP05-3	5.65	0.033	0.00	2.54
TSP05-4	2.65	0.004	-0.69	19.35
TSP05-5	0.00	0.00	-0.44	-
TSP05-6	0.00	0.00	0.00	7.22
TSP05-7	7.30	0.00	0.00	2.59
TSP05-8	2.60	0.00	-1.40	19.31
TSP05-9	0.00	0.00	-1.37	12.69
US01-39	1.80	0.00	-0.39	15.30
US01-40	0.00	0.00	0.00	9.74
US02-97	0.00	0.00	-1.15	5.33
US02-98	8.25	0.009	-0.37	8.69
US90-18	3.65	-0.006	-2.02	5.59
US93-15	11.60	-0.042	-0.94	6.64
US93-17	4.35	0.013	0.00	-
US99-2	-	0.065	-3.86	-

a-chlorophyll content, b-chlorophyll fluorescence, c-Leaf temperature, d-leaf relative water content

APPENDIX D

Primer	Motif	Putative Function
EST-SSR4	$(GGC)_9$	No hits found
EST-SSR9	$(TTC)_{20}$	Proteinase phosphatase 2c homolog
EST-SSR12	$(GA)_{15}$	Similar to UP/O81915 (O81915) T7123.21 protein
EST-SSR15	$(GA)_{31}$	Similar to UP/Q9LGV5 (Q9LGV5) ESTs D41739(S4522)
EST-SCA048	$(CA)_8$	Chloroplast phytoene synthase 1 [Zea mays]
EST-SSR50	$(GAG)_5$	Probable ethylene response protein
EST-SSR52	$(TA)_5$	PRMS-maize pathogenesis related protein PRMS precursor
EST-SSR80	$(CGC)_5$	Cystein protease component of protease inhibitor complex
EST-SSR81	$(AT)_{23}$	Chitin Inducible gibberilin responsive protein
EST-SSR023	$(AG)_{10}$	Ubiquitin C-terminal hydrolase
EST-SSR230	$(TA)_{12}$	Patatin like protein
EST-SSR668	$(CT)_{10}$	ATP synthase
EST-SSR924	$(CTCTCC)_5$	Auxin-independent growth promoter

APPENDIX E

Primer	Alleles	T(°C)	PIC	Size (bp)	Primer sequence (5'- 3')
SSR4	1	72.5	0*	260	CAACAATCGGGATGTCACTG (F) GATGCACAACAACAAGCACC (R)
SSR9	7	70.5	0.82	204-320	AAGAAAAGGAGGGCCAAAAA (F) GCCAGGCAAGAGGATAAAAA (R)
SSR12	5	59.5	0.75	230	AGAAGGAACGGTACCACGAC (F) TTGAAGTCGAGCACGATGAG (R)
SSR15	2	73.5	0.37	220	ATCCCAGAGCCCATCTCC (F) ATCTCCATACCTCCCCAGCA (R)
SCA 48	9	62.5	0.84	254	GCAACTCCGGCCTCTCCT (F) TTTCTGTTTTGCTCCTCCGTCTG (R)
SSR50	2	54.5	0.37	187	CTGCTGCTGTGTGCTGTAGG (F) CAACTTTTCGCCCTCCAAT (R)
SSR52	2	57.5	0.32	227	TGCACACGGACGTGTCTATAA (F) TGGTACAACTACGCCACCAA (R)
SSR80	6	58.5	0.81	229	GTTCCCACCGCTGTCATC (F) TACGAGCACGTGTCCAACTC (R)
SSR81	8	59.5	0.83	229	TTCTGCGTGGCACTGACTAC (F) ACAAAGGGCATCCTTTCTGA (R)
SSR023	2	56	0.36	100-150	AACATTTCGGCATTTGAAGC (F) GGTCTTTCTTGGGGATCTCTC (R)
SSR230	5	55	0.74	200-380	TTGTGCTGATGTTTCCTGCT(F) CAAGAGAAGATGCCATTAGCC (R)
SSR668	6	58	0.79	250-400	CAACAATTGTCGAAGCCTCTC (F) TTTGCTTACCCCCTGTTGAC (R)
SSR924	1	56	0*	200-300	CCGAGTGTCCTCATCGCAGAAC (F) CTCTAGTCTCTTCATAACCTCTC (R)

- Polymorphic information content could not be calculated for the primers, which produced only one polymorphic band

APPENDIX F

NAME	EST	Gene	Forward Primer
TRAP 119	MCSA119H06	AP2 domain containing	AAAATGGAATGCTCTGGT
TRAP 202	SCPRLB2027C02.g	DREB	TCA TTCCTCTTCCACTCC
TRAP 312	SCSBRZ3121G08.g	DREB	GAGTCACGGAGTCAGGAT
TRAP 313	SCEPRZ3132A02.g	DREB	TATTGCTTCCCCTTCTTT
TRAP 423	LEAF4_23_B07	AP2 domain containing	CCGTAGATTGCATTGTTG
Arbi 2			GACTGCGTACGAATTTGC

VITA

Name:	Vivek Sharma
Education:	B.Sc. (Hons) Agriculture, C.C.S.H.A.U Hisar, Haryana, India (2000) M.Sc. Biotechnology and Molecular Biology, C.C.S.H.A.U Hisar, Haryana, India (2003) Ph.D. Plant Breeding Texas A&M University College Station, TX (2009)
Contact Information:	Texas AgriLIFE Research 2415 E Highway 83 Weslaco, TX 78596 Phone: 956.969.5618 (O) vsharma@tamu.edu vivekcertain@gmail.com